



UNIVERSIDADE DE BRASÍLIA
Faculdade de Ciência da Informação
Curso de Graduação em Biblioteconomia

INDEXAÇÃO SOB UM ENFOQUE LINGUÍSTICO

Marianna Florência de Sousa Carvalho
Orientadora: Prof^ª. Dr^ª. Fernanda de Souza Monteiro

Brasília

2019

Marianna Florência de Sousa Carvalho

INDEXAÇÃO SOB UM ENFOQUE LINGUÍSTICO

Monografia apresentada como parte das exigências para obtenção do título de Bacharel em Biblioteconomia pela Faculdade de Ciência da Informação da Universidade de Brasília.

Orientadora: Prof^a. Dr^a. Fernanda de Souza Monteiro

Brasília

2019

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo (a) autor (a)

C331i Carvalho, Marianna Florência de Sousa
 Indexação sob um enfoque linguístico/ Marianna
Florência de Sousa Carvalho; orientadora Fernanda de Souza
Monteiro. -- Brasília, 2019.
 63 p.

Orientador: Fernanda de Souza Monterio

Monografia (Graduação - Biblioteconomia) -- Universidade
de Brasília, 2019

1. Indexação. 2. Linguística. 3. Análise conceitual. 4.
Tradução em Linguagem documentária. I. Monteiro, Fernanda
de Souza, orient. II. Título.



Título: Indexação sobre o enfoque linguístico.

Aluna: Marianna Florência Carvalho

Monografia apresentada à Faculdade de Ciência da Informação da Universidade de Brasília, como parte dos requisitos para obtenção do grau de Bacharel em Biblioteconomia.

Brasília, 06 de dezembro de 2019.

Fernanda de Souza Monteiro - Orientadora
Professora da Faculdade de Ciência da Informação (FCI/UnB)
Doutora em Ciência da Informação

Felipe Augusto Arakaki - Membro
Professor da Faculdade de Ciência da Informação (FCI/UnB)
Doutor em Ciência da Informação

Helena da Silva Guerra Vicente - Membro
Professora do Departamento de Linguística da Universidade de Brasília
(LIP/UnB)
Doutora em Linguística

AGRADECIMENTOS

Primeiramente, eu quero agradecer a Deus por estar comigo em todos os momentos e me dar sabedoria para realizar este trabalho. Em seguida, eu sou grata a meus pais e família pelo apoio, carinho e paciência que tiveram comigo. De forma semelhante agradeço aos meus amigos, colegas e conhecidos que me apoiaram e disseram que ia dar tudo certo.

Eu sou grata também a minha orientadora, Fernanda de Souza Monteiro, pela dedicação, correções e paciência comigo durante todo o processo. Agradeço a oportunidade de conhecer profissionais incríveis durante o curso de Biblioteconomia, como professores, palestrantes e bibliotecários inspiradores.

Por fim, sou grata por ter chegado até aqui e ter tido forças e determinação para concluir mais um curso. Eu sou grata pela minha vida e por estar aqui neste momento escrevendo, pois ainda não cheguei aonde quero estar, mas com certeza estou caminhando para chegar lá.

“Eis o meu segredo: só se vê bem com o coração. O essencial é invisível aos olhos. Os homens esqueceram essa verdade, mas tu não a deves esquecer. Tu te tornas eternamente responsável por aquilo que cativas.”

Antoine de Saint-Exupéry

RESUMO

A Biblioteconomia relaciona e utiliza de metodologias da Linguística, que auxiliam principalmente no processo de Indexação durante a realização das etapas de análise conceitual e tradução para linguagens documentárias. Entretanto, os alunos ingressos no curso de biblioteconomia não necessariamente tem um bom conhecimento da norma culta padrão da língua portuguesa. Portanto, objetivo do trabalho é identificar os aspectos linguísticos necessários para a formação de indexadores. Para alcançar o objetivo, realizou-se um levantamento bibliográfico, esclareceram-se os aspectos conceituais da Indexação e os aspectos conceituais da Linguística, e realizou-se a interseção das etapas de Indexação e conceitos linguísticos. Por fim, demonstra-se, na prática, como os conhecimentos linguísticos auxiliam os indexadores na indexação de um texto jornalístico destacando os conceitos relevantes nesse processo. Conclui-se que é importante o ensino de Linguística no curso de Biblioteconomia visando auxiliar os alunos nos diferentes níveis de análise linguísticos dos documentos e evitando erros no processo de indexação.

Palavras-chave: Indexação. Linguística. Análise conceitual. Tradução em Linguagem documentária.

ABSTRACT

Librarianship relates and uses linguistic methodologies, which mainly assist in the Indexing process during the conceptual analysis and translation steps for documentary languages. In other hand, the students who started course of Librarianship not necessarily has a good knowledge of the standard cultured pattern of the Portuguese language. Therefore, the objective of this work is to identify the linguistic aspects necessary for the development of indexers. To reach the objective, we made a bibliographic survey, the conceptual aspects of Indexing and the conceptual aspects of Linguistics were clarified, and the intersection of Indexing steps and linguistic concepts was established. Finally, we demonstrate in practice how linguistic skills assist indexers at indexing a journalistic text by highlighting the relevant concepts in this process. Our conclusion is that the teaching of Linguistics in the course of Librarianship is important, aiming to assist students in different levels of linguistic analysis of documents and avoiding errors in the indexing process.

Keywords: Indexing. Linguistics. Conceptual analysis. Translation in Documentary language.

SUMÁRIO

1. INTRODUÇÃO	10
2. METODOLOGIA	14
3. INDEXAÇÃO - ASPECTOS CONCEITUAIS	18
3.1. EVOLUÇÃO CONCEITUAL	18
3.2. TEORIA DE CONCEITO e TERMO	20
3.3. ASPECTOS QUALITATIVOS DA INDEXAÇÃO	21
3.3.1. Exaustividade	21
3.3.2. Especificidade	21
3.3.3. Coerência	22
4. LINGÜÍSTICA - ASPECTOS CONCEITUAIS	23
4.1. BREVE HISTÓRICO	23
4.2. CONCEITOS	25
4.2.1. Linguagem natural	25
4.2.2. Linguagem Documentária	26
4.2.3. Morfologia	27
4.2.4. Sintaxe	28
4.2.5. Semântica	29
4.2.6. Pragmática	32
4.2.7. Terminologia	33
5. ETAPAS DA INDEXAÇÃO E LINGÜÍSTICA	37
5.1. ANÁLISE CONCEITUAL.	37
5.2. TRADUÇÃO EM LINGUAGEM DOCUMENTÁRIA	41
6. PRÁTICA DE INDEXAÇÃO	45
7. CONCLUSÃO	56
8. REFERÊNCIAS	59
ANEXO 1 – Reprodução integral do texto jornalístico analisado	62

1. INTRODUÇÃO

A Ciência da Informação (CI) tem como uma de suas incumbências organizar e representar a informação para a recuperação e uso pela sociedade. De natureza interdisciplinar, ela relaciona-se com outras ciências e toma-lhes emprestados princípios, práticas e teorias, a título de exemplo o que ocorre com a Linguística. Um dos componentes da CI é a Biblioteconomia. No decorrer do curso de Biblioteconomia, ministrado na Universidade de Brasília (UnB), as disciplinas Análise da Informação, Classificação e Indexação destacaram as influências e princípios da Linguística utilizados nos processos de representação temática da informação.

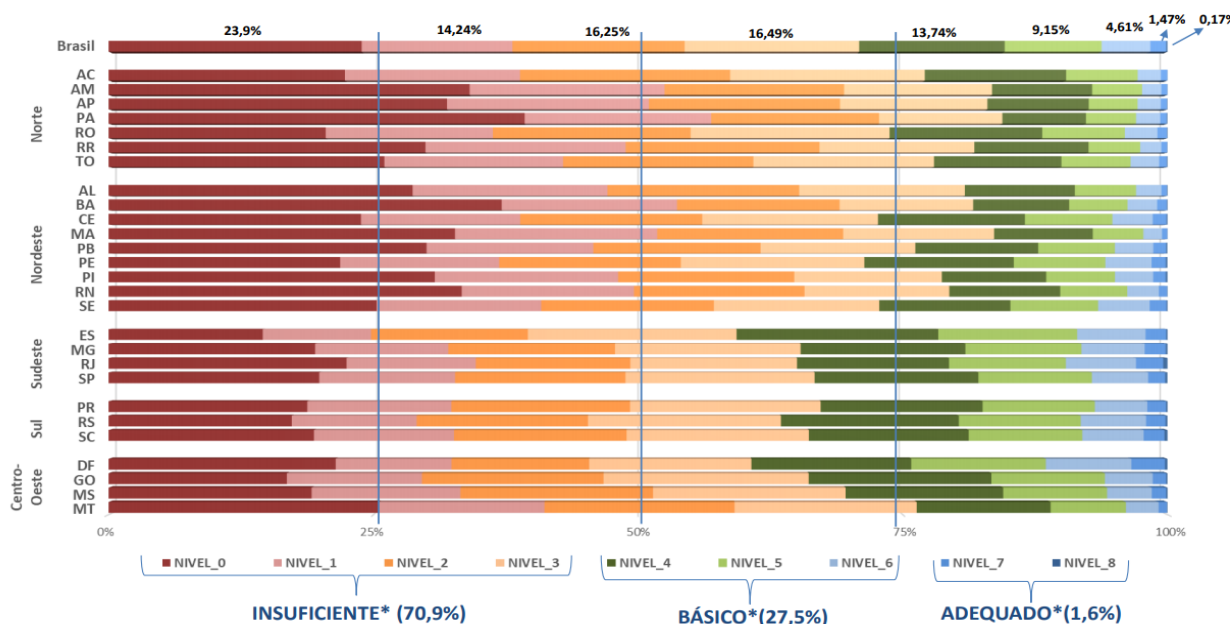
Com a base nos conhecimentos da área de Linguística que o curso de Letras inglês licenciatura proporcionou à autora, foram realizadas conexões entre os conceitos da Ciência da Informação e Linguística que ampliaram a compreensão de conceitos, como: *termo e conceito*, *relações hierárquicas partitivas*, *relações hierárquicas de gênero e espécie*, entre outros. Porém os alunos ingressos em Biblioteconomia não necessariamente possuem essa base.

De acordo com os resultados do Sistema de Avaliação da Educação Básica (Saeb), um processo de avaliação da qualidade da educação praticada no país, por meio de testes e questionários, que refletem os níveis de aprendizagem de Língua Portuguesa e de Matemática demonstrados pelo conjunto de estudantes avaliados¹, no ano de 2017, apenas 1,6% dos estudantes do país que participaram do Saeb apresentaram aprendizagem adequada de Língua Portuguesa segundo o MEC.

¹ O exame é realizado periodicamente pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep).

SISTEMA DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA | Edição 2017

Distribuição (%) dos estudantes nos níveis da Escala de Proficiência Ensino Médio – Língua Portuguesa



* Interpretação dada pela Secretaria de Educação Básica do Ministério da Educação.

INEP

MINISTÉRIO DA EDUCAÇÃO

Sistema de Avaliação da Educação Básica. Edição 2017. Escala de Proficiência Ensino Médio – Língua Portuguesa. Fonte: http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=94181-saeb-2017-versao-ministro-revfinal-1&category_slug=agosto-2018-pdf&Itemid=30192.

Tendo em mente a realidade sobre a aprendizagem de português nas escolas, depreende-se uma das razões de conceitos linguísticos advindos de níveis de análise linguística, tais como Sintaxe, Semântica, Morfologia, Pragmática e Terminologia não estarem claros na mente dos alunos, pois alguns desses conceitos só são estudados mais especificamente ao nível de graduação e outros podem ser aprendidos de forma precária no Ensino Básico e no Ensino Médio, suscitando lacunas e níveis inadequados de aprendizado. Assim, o aluno ingresso no curso de Biblioteconomia não necessariamente tem uma boa base para o uso da norma padrão da Língua Portuguesa, tornando difícil o uso desta para as metodologias da Ciência da Informação.

Com a base desigual de aprendizagem e uso da norma padrão Língua Portuguesa entre os estudantes, tentar suprir o conhecimento nos diferentes níveis de análise linguística durante a formação do bibliotecário é determinante para a disseminação da informação documentária, pois a Linguística está relacionada à representação, em especial, à Classificação, Indexação, Análise da informação, elaboração de resumos, etc.

Uma das áreas de atuação do bibliotecário é a indexação de documentos, sendo a influência da linguística nas etapas de indexação o foco do presente trabalho. Portanto, um bom indexador tem que ter domínio da linguagem utilizada no documento para realizar a leitura documentária, levantar os conceitos abordados e traduzir em termos de Indexação e resumos, possibilitando a recuperação precisa do documento pelo usuário.

A partir dos destaques acima e da análise dos conteúdos de ementas das disciplinas Classificação, Análise da Informação e Indexação, da área da Ciência da Informação, e as disciplinas Introdução à Linguística, Sintaxe do Português Contemporâneo, Morfologia do Português e Leitura e Produção de Textos, da área da Linguística, no âmbito dos cursos ministrados na Universidade de Brasília, surgiu o questionamento de como ocorrem as relações da Linguística com a Biblioteconomia e levantaram-se os pontos de interseção de conteúdo entre elas, destacando os níveis de análise linguística: Morfologia, Sintaxe, Semântica, Pragmática e Terminologia. Estabeleceu-se o objetivo geral de identificar e demonstrar quais conhecimentos necessários da área de Linguística precisam ser abordadas no curso de Biblioteconomia, visando evitar erros de Indexação e falhas na recuperação da informação, além de suprir lacunas linguísticas na formação de indexadores. Como objetivos específicos, pretende-se:

1. Apresentar os aspectos conceituais da Indexação;
2. Apresentar os aspectos conceituais da Linguística;
3. Realizar a interseção das etapas de Indexação e conceitos linguísticos.

Cada objetivo específico foi abordado em um capítulo. O capítulo três, intitulado “Indexação: aspectos conceituais”, aborda a evolução das correntes teóricas na literatura especializada, as etapas de Indexação, definição de termo e conceito, e alguns parâmetros de qualidade.

O capítulo quatro, intitulado “Linguística: aspectos conceituais”, aborda a origem dessa ciência e aspectos conceituais relevantes para a prática de Indexação de forma pontual. O capítulo cinco, intitulado “Etapas da Indexação e a Linguística”, aborda a análise das duas etapas da Indexação de acordo com a definição da UNISIST:

A indexação é vista como a ação de descrever e identificar um documento de acordo com seu assunto. [...] Durante a indexação, os conceitos são extraídos do documento através de um processo de análise, e então traduzidos para os termos de instrumento de indexação (tais como tesouros, listas de cabeçalhos de assunto, esquemas de classificação, etc.) (UNISIST, 1981, p. 84)

Abordaremos a análise conceitual e tradução em linguagens documentárias, como etapas da indexação, com os conceitos da Linguística pertinentes a essas etapas. O capítulo seis, intitulado “Prática de Indexação”, demonstra como seria, na prática, a indexação de um artigo de jornal salientando a análise linguística no processo através de exemplos extraídos do texto. O capítulo sete, intitulado “Conclusão”, traz os resultados do trabalho.

Portanto, este estudo visa demonstrar quais conhecimentos necessários da área de Linguística precisam ser abordados no curso de Biblioteconomia, com o objetivo de evitar erros de Indexação e falhas na recuperação da informação, além de suprir lacunas relacionadas ao conhecimento dos níveis de análise linguística.

2. METODOLOGIA

A abordagem deste estudo é de base fenomenológica, pois foca nas relações entre Linguística e as etapas de Indexação esclarecendo essas relações. Utiliza-se do método de técnica comparativa para explicar o fenômeno das relações e analisar o dado concreto, ressaltando as similaridades entre eles, e deduzindo deste os conhecimentos necessários da área de Linguística para cientistas da informação.

Esse método se baseia na Etnometodologia, de Harold Garfinkel, em que “os objetos sociais são criados a cada ocasião situada, por pessoas que cooperam a fim de cumprir critérios constitutivos, e de que elas continuam em mudança a cada novo movimento” (RAWLS, 2018, p. 445). Os objetos e suas relações são estudados ao longo do tempo com o envolvimento e a inclusão do observador no processo. As etapas metodológicas para alcançar o objetivo da pesquisa de identificar os aspectos linguísticos necessários para a formação de indexadores são:

1. Analisar os conteúdos abordados em ementas das disciplinas Classificação, Análise da informação e Indexação, da área da Ciência da Informação, e as disciplinas Introdução à Linguística, Sintaxe do Português Contemporâneo, Morfologia do Português e Leitura e Produção de Textos, da área da Linguística. Essas disciplinas foram selecionadas no currículo disponível do site *matriculaweb.unb.br*, dos cursos de Biblioteconomia e Letras Português da Universidade de Brasília, devido a corresponderem à realidade contextual dos estudantes de Biblioteconomia da Universidade de Brasília, e por serem disciplinas que abordam os aspectos linguísticos metodológicos em suas respectivas áreas. As ementas estão explicitadas a seguir:

Em *análise da informação*: a Conceituação, fundamentação teórica, objeto e função da análise da informação. Contextualidade da informação no ciclo documentário. Subsídios interdisciplinares para a análise da informação: Linguística, Semântica, Semiótica. Terminologia e Lógica. Métodos e técnicas da análise da informação. Métodos e técnicas de análise, síntese e representação da informação. O resumo documentário e a indexação.

Em *Classificação*: a Função e valor do pensamento classificatório. Conceitos fundamentais. Origem e evolução dos sistemas de classificação. Sistemas de classificação e linguagens documentais. Macro e micro estruturados sistemas de classificação bibliográficas e das linguagens

documentais. Representação documentária por meio de classificação bibliográficas. Classificações bibliográficas de caráter enciclopédico. Classificações bibliográficas especializadas.

Em *Indexação: a Conceituação*, fundamentos teóricos, características e funções da indexação. Questões epistemológicas e metodológicas da indexação. Tipologia da indexação e dos índices. Instrumentos de métodos de controle terminológico. As linguagens documentárias utilizadas na indexação. Indexação automática.

Em *Introdução à Linguística*: O estudo científico da linguagem: noções básicas. Língua e cultura. Gramática tradicional, Linguística Formal e Linguística Funcional. Variação linguística. Língua Padrão. Atitudes e preconceitos linguísticos. Aquisição da língua. Competência comunicativa.

Em *Sintaxe do Português Contemporâneo I*: Estruturas sintáticas básicas do português contemporâneo: Estruturação da oração e do período. Sintaxe de colocação de concordância e de regência. Sintaxe e pontuação. Emprego das classes de palavras. Processos de análise sintática. Sintaxe e Estilo.

Em *Morfologia do Português*: conceitos e procedimentos linguísticos: o morfema e a palavra; depreensão e classificação de morfemas; processos morfológicos; classes de palavras. Análise morfológica do português: flexão; processo de criação lexical; classes gramaticais.

Em *Leitura e Produção de textos*: a leitura ativa, analítica e crítica de textos. Planejamento e produção de resumos, resenhas críticas e textos dissertativos - argumentativos.

2. Levantar os pontos de interseção entre os conteúdos das disciplinas que nortearão o levantamento bibliográfico. Analisaram-se os pontos convergentes entre as ementas, ou que podem colaborar entre si: a Competência comunicativa que leva em conta não apenas os conhecimentos normativos da gramática da língua Portuguesa, mas contextuais de comunicação que são estudados no nível de análise da Pragmática; os Sistemas de classificação que utiliza a língua e a classificação de palavras em linguagens documentárias que são estabelecidos através de campos semânticos que são estudados no nível de análise da Semântica.

A Metodologia de Indexação que abrange a análise, síntese e representação da informação em palavras-chave extraídas de Instrumentos de métodos de controle terminológico que são estudados no nível da análise da Terminologia; Processo de criação lexical que é estudado no nível da análise da Morfologia; Processos de análise sintática, leitura ativa, analítica e crítica de textos, planejamento e produção de resumos que são estudados a nível da análise da Sintaxe; e outros subsídios interdisciplinares para a análise da informação que são advindos da Linguística. Definiram-se os seguintes termos de busca: Indexação, Linguística, Morfologia, Sintaxe, Semântica, Pragmática e Terminologia;

3. Levantar documentos utilizando os pontos de interseção das áreas como termo de busca. Os instrumentos que serão utilizados são fontes bibliográficas já publicadas em diversas Bases de Dados: Ciência da Informação, ENANCIB - Encontro Nacional de Pesquisa em Ciência da Informação, Brapci, Informação & Sociedade: Estudos (I&S), RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação, TransInformação, e Brazilian Journal of Information Science: research trends que são relevantes na área da Ciência da Informação, pois se visa observar como o assunto é tratado na CI em específico, de acordo com o olhar da biblioteconomia, não sendo utilizado Bases de Dados da área da Linguística.

Além da pesquisa no catálogo geral da Biblioteca Central dos Estudantes (BCE) da Universidade de Brasília, que é pertinente à realidade contextual dos estudantes de Biblioteconomia da Universidade de Brasília, a fim de complementar os assuntos linguísticos quando eles se apresentarem insuficientes nas fontes de informação da CI.

Após o levantamento, observa-se que as bases de dados específicas da área da Ciência da Informação contêm mais textos que abordam contribuições de conceitos advindos dos níveis semântico, pragmático e terminológico do que dos níveis morfológico e sintático. Denota-se a necessidade de mais pesquisas sobre as contribuições e interligações entre as áreas.

De forma semelhante, percebe-se que o catálogo geral da BCE, por não ser específico da CI, contém mais itens dos referidos termos da Linguística do que as bases específicas da Ciência da Informação.

4. Realizar uma leitura documentária do título, resumo, e título de capítulos, e selecionar textos pertinentes à pesquisa;

5. Ler os documentos selecionados de forma completa e conceituar os aspectos evolutivos da Indexação, etapas da Indexação, Termo e Conceito, e aspectos qualitativos como: Exaustividade, Especificidade e Coerência;

6. Ler os itens selecionados e descrever o histórico da Linguística. Além de explicar e exemplificar alguns conceitos como: Linguagem natural, Linguagem documentária, Morfologia, Sintaxe, Semântica, Pragmática e Terminologia;

7. Relacionar as etapas de Indexação com os conceitos da Linguística, destacando a contribuição dos níveis de análise linguística denominados Morfologia, Sintaxe, Semântica, Pragmática e Terminologia nas etapas por meio de associação dos conceitos;

8. Realizar uma prática de Indexação de um texto jornalístico, destacando, em quadros, as análises linguísticas que contribuem para a prática, utilizando exemplos extraídos do texto.

Essa pesquisa possui caráter teórico e natureza básica, pois objetiva gerar conhecimentos úteis para o avanço da ciência. Sob o ponto de vista de seus objetivos, é considerada exploratória e tem como finalidade proporcionar mais informações sobre o assunto que vamos investigar, possibilitando sua definição e seu delineamento.

3. INDEXAÇÃO - ASPECTOS CONCEITUAIS

3. 1. EVOLUÇÃO CONCEITUAL

Os primeiros rastros da Indexação datam do século II A.C. com a produção de ‘resumos’ nas tábuas de argila produzidas na Mesopotâmia. Avançando na história, a Indexação começa com o enfoque de elaboração de índices que, neste primeiro momento, tem o sentido tradicional, segundo Robredo (1994 apud SILVA; FUJITA, 2004), e consiste em uma lista alfabética ou sistemática de tópicos que se ligam a uma localização no documento. Na biblioteca de Alexandria, usava-se a classificação de Calímaco, em que o catálogo era organizado por ordem alfabética de autores e subordinados a assuntos mais gerais.

No século II, surgem os primeiros guias para cabeçalhos de capítulos, tábuas de matérias ou sumários, cabeçalhos nas margens dos parágrafos, cabeçalhos descritos no alto da página, intitulado *De Libris Propriis Libris* compilado por Cláudio Galeno. Esses sumários eram encontrados em textos bíblicos e foram mencionados por Cassiodoro em seu trabalho:

²Estes sumários (tituli) são mencionados algumas vezes por Cassiodoro em seu *Institutiones**, o qual ele preencheu com cabeçalhos no começo de cada livro para ajudar na busca de informações neles. Enquanto isso parece, de alguma forma, longe da indexação, como nós conhecemos, isso permitiu uma busca rápida por data e capacitou Cassiodoro a fazer referência cruzada em seu texto. (WITTY, 1973, p. 194 - tradução do autor)

No século V, é realizado um trabalho de índice alfabético de assunto sobre provérbios gregos de tópicos teológicos intitulado *Apothegmata*, uma obra anônima. No século XIV, eram elaborados inventários e catálogos, com lista de conteúdo e resumos, de livros existente em mosteiros. Os copistas escreviam, à margem do livro, os pontos principais dos textos, mas essa interpretação diferia entre eles afetando a qualidade desse princípio de análise conceitual.

Em um segundo momento da história, o índice tem o sentido amplo, expresso em Robredo (1994 apud SILVA; FUJITA, 2004), de chave que representa assuntos, tópicos ou conceitos, e que

²These summaries (tituli) are mentioned a number of times by Cassiodorus in his *Institutiones**, which he furnished with such headings at the beginning of each book to aid in finding information contained therein. While this seems somewhat far from indexing, as we know it, it did permit easier searching of data and enable Cassiodorus to cross-reference his text (WITTY, 1973, p. 194).

pode servir como critério para localizar nos documentos os assuntos. No século XVII, com a compilação da primeira concordância completa da Bíblia inglesa por Alexandre Cruden, houve a necessidade de criar um índice que vinculava as citações a sua localização no texto. No mesmo período, começava os tratamentos de periódicos semanais com a edição de *Le Journal des Sçavans* publicado em Paris, que traziam resumos de trabalhos científicos, filosóficos e artísticos.

No século XVIII, Johnson estabeleceu os termos de entradas e as passagens a serem indexadas indexou na língua inglesa, segundo Silva e Fujita (2004, p. 139). Cruden e Johnson são os primeiros a se preocuparem com padrões de clareza e consistência na indexação. Na Alemanha surge o conceito de palavra-chave. Nos séculos XIX e XX, devido à explosão documental, a Indexação evoluiu para índices de vários volumes, cooperativos e em nível internacional. No século XXI, a Indexação está a caminho de se tornar automática com o auxílio de várias áreas do conhecimento como Linguística, Engenharia, entre outras.

Como pode-se ver, a Indexação surgiu com a elaboração de índices, porém atualmente ela está mais relacionada a análise de conteúdo. O conceito de Indexação permeia por duas correntes teóricas na literatura: a francesa e a inglesa. Na francesa, a Indexação “é vista como uma operação de representação documentária com a finalidade pragmática de Recuperação da informação” (SILVA; FUJITA, 2004, p. 136). Nessa corrente, a Indexação é o processo final da análise documentária em que se faz uso de linguagens documentárias para produzir produtos documentários.

Já na corrente inglesa, a Indexação é “a própria análise documentária, compostas das mesmas etapas operacionais com o objetivo de representação do conteúdo informacional de documentos para a elaboração de índices.” (SILVA; FUJITA, 2004, p. 136). Nessa corrente, a Indexação e análise documentária são os mesmos processos, sendo a análise a primeira parte da Indexação. Sendo assim, define-se Indexação, segundo a UNISIST, em:

A indexação é vista como a ação de descrever e identificar um documento de acordo com seu assunto. [...] Durante a indexação, os conceitos são extraídos do documento através de um processo de análise, e então traduzidos para os termos de instrumento de indexação (tais como tesouros, listas de cabeçalhos de assunto, esquemas de classificação, etc.) Decidida a indexação, os conceitos são registrados como “dados de informação” e são organizados de modo a permitir um fácil acesso na recuperação da informação. Estes registros podem aparecer nas mais variadas formas, ex. índice de livros, índices para catálogos e bibliografias, catálogos em fichas, etc. (UNISIST, 1981, p. 84)

De forma semelhante, na Associação Brasileira de Normas Técnicas, a NBR 12676 (1992, p. 2) define Indexação como “ato de identificar e descrever o conteúdo de um documento com termos representativos dos seus assuntos e que constituem uma linguagem de indexação”. A autora, para o presente trabalho, adota a concepção de Indexação com duas etapas segundo Lancaster (2004, p. 8-9): “A indexação de assuntos envolve duas etapas principais: 1. Análise conceitual, e 2. Tradução”; essas etapas foram descritas com mais detalhes no capítulo 6. A seguir, definem-se Termo e Conceito, que compõem as linguagens documentárias utilizadas na Indexação e os parâmetros de Indexação.

3.2. TEORIA DE CONCEITO e TERMO

Na Indexação, como definido pela UNISIST (1981), extraem-se conceitos dos documentos que serão traduzidos em *termos*, e esses *termos* são extraídos de linguagens documentárias, como tesouros, visando a representação e recuperação desta informação. Para compreender esse processo, veremos a concepção de *conceito* e *termo* nas visões de Dahlberg e Hjørland.

O *conceito*, segundo Dahlberg, é “[...]a compilação de enunciados verdadeiros sobre determinado objeto, fixada por um símbolo linguístico[...]” (DAHLBERG, 1978b, p. 102, apud MELO & BRACHËR, 2015, p. 71). Essa visão de *conceito* é mais rígida não abrangendo a dinâmica e as diferentes realidades que um conceito pode descrever, como por exemplo, o conceito de “lar”, que é único e relativo ao emissor desta palavra. Essa dinamicidade do *conceito* é abordada na definição de Hjørland, em que *conceito* “[...]são significados dinamicamente construídos e negociados que classificam o mundo de acordo com os interesses e teorias[...]” (MELO & BRACHËR, 2015, p. 73). Salienta-se que *conceito* é diferente de palavra, pois segundo Hjørland “[...]um conceito expressa apenas um significado; enquanto palavras diferentes podem ter o mesmo significado e palavras similares podem ter diferentes significados[...]” (MELO & BRACHËR, 2015, p. 70).

Para representar os *conceitos* utilizamos símbolos linguísticos denominados *termos*. Na visão de Dahlberg, o *termo* é “[...]o componente que, convenientemente, sintetiza e representa um conceito com o propósito de designá-lo e comunicá-lo[...]” (MELO & BRACHËR, 2015, p. 71). Entretanto, a língua é um instrumento em constante mudança, e sobre a dinamicidade do *termo* o autor Hjørland discorre que “[...]o significado de um termo não pode ser algo perene, fixado

definitivamente, pois os conceitos são dinâmicos e evoluem segundo a construção social do conhecimento[...]” (MELO & BRACHËR, 2015, p. 73).

Na Ciência da Informação, padronizar a descrição de documentos auxilia a juntar assuntos relacionados e a busca de documentos pelos usuários. Então, por mais que um *conceito* possa mudar com o tempo e contexto, eles são vinculados a termos e especificados através de definições em instrumentos de indexação como os tesauros. Porém, é necessário a atualização desses instrumentos sempre que a mudança conceitual for pertinente.

3.3. ASPECTOS QUALITATIVOS DA INDEXAÇÃO

Nas linguagens documentárias, fixa-se o *conceito* a um *termo* a fim de padronizar a descrição de documentos com o auxílio de tesauros e outros instrumentos. Para que se tenha um padrão, deve-se levar em conta alguns aspectos qualitativos da indexação, como exaustividade, especificidade e coerência.

3.3.1. Exaustividade

A exaustividade é o emprego de termos de indexação para abranger de forma completa o conteúdo do documento, e está relacionada à política institucional de indexação que pode ser exaustiva ou seletiva. Em Lancaster (2004) observa-se que:

A indexação exaustiva implica o emprego de termos em um número suficiente para abranger o conteúdo temático do documento de modo bastante completo. A indexação seletiva, por outro lado, implica o emprego de uma quantidade muito menor de termos, a fim de abranger somente o conteúdo temático principal do documento. (LANCASTER, 2004, p.27)

A mais exaustiva geralmente é mais onerosa para o sistema e menos precisa do que a seletiva. Quanto mais termos forem usados, mais vezes o documento será recuperado no processo de busca. Contudo, pode ser indexado um termo que não corresponde precisamente ao que o usuário está buscando, pois, o assunto indexado pode não ser o assunto principal do documento, mas apenas uma citação.

3.3.2. Especificidade

Esse aspecto está relacionado ao nível de detalhamento da descrição, como definido em Lancaster (2004, p. 34) “ O princípio que, isoladamente, é o mais importante da indexação de

assuntos, e que remonta a Cutter (1876), é aquele segundo o qual um tópico deve ser indexado sob o termo mais específico que o abranja completamente”.

O indexador pode utilizar o termo mais geral ou o termo mais específico em uma linguagem documentária para descrever um item, dependendo da necessidade de detalhamento do sistema de informação ou da necessidade do usuário. Observa-se que o indexador terá que decidir sobre a exaustividade ou a especificidade, pois segundo Fujita (2012) a especificidade leva a uma recuperação da informação com alta precisão e baixa revocação, enquanto a exaustividade gera uma recuperação com baixa precisão e alta revocação.

3.3.3. Coerência

Por ser um processo subjetivo da indexação pode ser de dois tipos, de acordo com Lancaster (2004, p. 68): “A *coerência Inter indexadores* refere-se à concordância entre indexadores, enquanto a *coerência intra-indexador* refere-se à extensão com que um indexador é coerente consigo mesmo”. Esse aspecto é importante para a padronização idealizada por cientistas da informação. Sem coerência, documentos de assuntos semelhantes podem ficar distantes nas estantes, ou não serem revocados quando são pertinentes ao assunto da busca. Para auxiliar a coerência, os instrumentos de indexação devem ser usados, como salienta Slamecka e Jacoby (1963, p. 30, *apud* LANCASTER, 2004, p. 75):

A coerência interindexadores melhora significativamente com a utilização de instrumentos de indexação prescritivos que contenham um mínimo de relações semânticas variáveis entre os termos. O emprego de instrumentos de indexação que ampliem a liberdade semântica do indexador, no que concerne à escolha de termos, é prejudicial à confiabilidade da indexação. A qualidade da indexação tem muito a ganhar com vocabulários que formalizem as relações de modo a prescrever uniforme e invariavelmente a escolha dos termos de indexação.

Portanto, para uma ‘boa’ indexação é importante considerar os aspectos qualitativos da indexação quanto ao grau de especificidade, exaustividade e o uso de instrumentos que auxiliem na coerência intra e entre indexadores. A fim de relacionar os métodos da Linguística que podem ajudar para uma ‘boa’ Indexação, serão abordados, no próximo capítulo, um breve histórico da Linguística e alguns conceitos que auxiliam a atividade de Indexação.

4. LINGÜÍSTICA - ASPECTOS CONCEITUAIS

4.1. BREVE HISTÓRICO

A Linguística é definida como “o estudo científico da lingua(gem)”, segundo Weedwood (2002, p. 9). Ela se tornou uma ciência em 1916 a partir da publicação póstuma do *Cours de Linguistique Générale*, de Ferdinand de Saussure. A teoria Estruturalista de Saussure tem como base duas dicotomias: a) Sistema Linguístico, que se refere a padrões de formação de uma língua, e b) Comportamento Linguístico, que se refere a enunciados reais dos falantes; e a) Forma, ou mais conhecido como Significante, que se refere à parte física da língua, e b) Substância, ou mais conhecido como Significado, que é o objeto em si desenvolvido na mente do falante.

Em 1930, nasce a teoria dos Campos Semânticos. Essa teoria, segundo Jost Trier, afirmava que nenhuma palavra é isolada no campo linguístico do falante, entretanto, ela participa de um conjunto lexical agrupado semanticamente. Esse campo linguístico é ligado por nós semânticos que constituem as palavras, ou seja, “Toda palavra é formada de um nó semântico mais ou menos denso, ou mais ou menos volumoso, cercado por uma auréola de associações secundárias, afetivas ou sociais” (GUIRAUD, 1978, apud CINTRA, 1983, p.12).

Na mesma época, surge a teoria da Análise Distribucional, que perdura até a década de 1950, e tem por base a teoria psicológica behaviorista. Ela analisa que a fala não é um processo aleatório, e o comportamento humano empregado na fala é passível de uma ordem para formar palavras e combinação de palavras para formar frases (análise sintagmática), além da relação das palavras presentes na frase com as palavras não presentes, mas pertencentes à mesma classe (análise paradigmática). Esses processos da fala são elucidados a seguir:

Na sequência "Gostaria de comprar uma fazenda", a unidade "comprar", por exemplo, ao mesmo tempo em que se encontra em relação paradigmática com "vender", "entregar", "olhar" e tantas outras unidades, também mantém relações sintagmáticas com "gostaria", "de", "uma" e "fazenda". [...]Esses fatos nos permitem compreender melhor o porquê da língua ser um sistema, uma estrutura, e não uma mera reunião de elementos. Adotando uma perspectiva estruturalista, podemos afirmar, então, que o que permite o funcionamento da língua é o sistema de valores constituído pelas associações, combinações e exclusões verificadas entre as unidades linguísticas. (MARTELOTTA, 2013, p.122).

Na década de 50, observou-se que a criatividade da linguagem, ou seja, a capacidade de criar ou decodificar enunciados nunca produzidos ou ouvidos pelo falante, mas que constituem

construções possíveis e previstas na língua, não estava sendo considerada pelos modelos estruturais vigentes. Em 1957, Noam Chomsky formulou a teoria Gerativo-Transformacional, que se distanciava do Estruturalismo e do Behaviorismo, e se preocupava com a criatividade linguística e a relação linguagem-pensamento levando em conta a diferença entre os níveis superficial (a estrutura visível da linguagem) e profundo (o seu significado no campo semântico) da estrutura gramatical.

Em sua dicotomia, ele dividiu: a) o Conhecimento da estrutura de uma língua que estimula a faculdade inata da linguagem, a faculdade é definida como uma gramática universal, ou estado cognitivo inicial, que o ser humano nasce com ela em sua mente e que é ativada ao receber um estímulo externo, e b) a Capacidade de produção de infinitas frases, ou seja, o uso efetivo da língua em situações reais de enunciados novos para o falante, mas possível e previsto na língua. Essa dicotomia ficou mais conhecida como “Àquele conhecimento ele se referiu como *competência* (*competence*) e ao uso como *desempenho* (*performance*)” (WEEDWOOD, 2002, p. 133).

Em 1960, na Escola de Praga, surge a *Linguística Sistemica* da teoria de Halliday, nela a gramática é analisada através de uma rede de sistemas, dando atenção a aspectos semânticos e pragmáticos. Nessa mesma escola, a noção de marcação foi desenvolvida, ou seja, presença ou ausência de um traço distintivo entre palavras. Primeiramente, entre estudos fonológicos e, depois, em estudos morfológicos e sintáticos. Essa marcação pode ser vista, por exemplo, na flexão de gênero do português: gato (masculino) e gata (feminino).³

Em 1968, a partir da revolução Chomskyana da teoria Gerativo-Transformacional, Charles Fillmore apresentou uma nova teoria denominada Gramática de Casos. Em sua teoria, as infinitas frases possíveis dos falantes contêm verbos, adjetivos, substantivos, artigos ligados com uma ou mais palavras desempenhando funções sintáticas, porém com uma função semântica (categorias) que varia de acordo com o caso:

A gramática de casos admite o mesmo princípio básico da gramática gerativo-transformacional: sobre a EP⁴ aplicam-se regras de transformação e se obtém a frase superficial. Entre as transformações uma dá origem ao objeto, outra ao sujeito, etc. E o sujeito então, sendo uma questão superficial, pode ser:

³ Ressalta-se que quanto a flexão de gênero há correntes divergentes, como a corrente do Mattoso Câmara Jr. Em que o “o” não é uma flexão de gênero masculino, mas sim neutro, e o “a” é uma flexão de gênero feminino.

⁴ EP refere-se a estrutura profunda, ou seja, a semântica da estrutura gramatical.

agente: Ele bate na bola.

paciente: Ele recebeu um golpe.

beneficiário: Ele recebeu um presente. (CINTRA, 1983, p. 20-21)

No século XX, houve uma atenção para o contexto da linguagem com a atenção para a Pragmática, que é definida, segundo Weedwood, como “[...]os fatores que regem nossas escolhas linguísticas na interação social e os efeitos de nossas escolhas sobre as outras pessoas.” (WEEDWOOD, 2002, p. 144). A Linguística vem evoluindo até hoje, e ainda necessita de sistematização de alguns conceitos pertencentes a outros níveis de análise linguística, tais como a Análise do Discurso, a Linguística Forense, dentre outros.

4.2. CONCEITOS

Partindo dos pontos convergentes levantados através da análise das ementas das disciplinas tanto da área de Biblioteconomia quanto de Linguística, o recorte feito a seguir no universo da Linguística visa os aspectos que são úteis aos profissionais da informação para o desempenho de sua atividade profissional, como já abordado desde 1983 por Baranow.

Os aspectos visam a linguagem escrita e sua interpretação pelos profissionais da Ciência da Informação, delimitado em: noções de Linguagem natural e Linguagem documentária, os diferentes níveis de análise “[...]a morfologia, que é o estudo das construções das palavras; a sintaxe, que é o estudo de como as palavras podem ser combinadas em sentenças; e a semântica, que é o estudo do significado das palavras e das sentenças[...]” (CANÇADO, 2008, p. 15), além do contexto e uso da linguagem, ou seja, a sua Pragmática, e sua representação com a Terminologia.

4.2.1. Linguagem natural

A Linguagem natural (LN) é a linguagem usada pelos falantes de uma língua. É a linguagem empregada pelo próprio autor, ou a mais acessível e próxima da linguagem do usuário da informação, como salienta Assis:

Essa linguagem se caracteriza por ser a utilizada comumente pelo usuário possibilitando a este mais liberdade no momento da busca, sendo possível utilizar qualquer termo que julga ser necessário para realizar a pesquisa. Porém deve-se ter atenção quanto ao uso, pois é um instrumento no qual as palavras podem apresentar vários significados. (ASSIS, 2010, p. 16)

A Linguagem natural é passível de mudança de significado feita pelo usuário no ato da fala sendo crucial o contexto comunicacional, pois a troca de um signo por outro pode acarretar em alteração na comunicação. A Linguagem natural é passível de falhas na comunicação, ou de erros gramaticais, não sendo a ideal para representação de informações. Ela se caracteriza como uma possível dificuldade para indexadores nos casos de polissemia, sinonímia, homonímia e antonímia que foram abordados na seção 4.2.5. Semântica.

4.2.2. Linguagem Documentária

A Linguagem documentária (LD), do ponto de vista semântico, vincula o Significante ao seu Significado, o que é fundamental para a recuperação de documentos na CI ao proporcionar uma maior precisão descritiva dos itens pertinentes ao assunto procurado pelo usuário da informação. Elas são construídas com o propósito de sobrepor a dificuldade de descrever com precisão os diferentes documentos. De acordo com Sayão (1996 apud ASSIS, 2010, p. 20), as linguagens documentárias:

São linguagens artificiais geralmente derivadas da linguagem natural. Essas linguagens são chamadas artificiais no sentido em que não resultam de processo evolutivo e necessitam de regras explícitas para seu uso. Normalmente as linguagens documentárias estabelecem uma relação unívoca entre o termo e o conceito, isto é, entre o significante e o significado. Cada termo corresponde a um conceito do sistema de conceitos da área específica com que está trabalhando. A construção dessas linguagens é um processo complexo e longo.

A LD desempenha a função de representação da informação documentária, atuando no controle do vocabulário de Indexação, e “[...]é utilizada no momento do tratamento intelectual dos documentos, isto é, no momento das operações de entrada no subsistema de armazenamento e de pesquisa da informação e no momento das operações de saída e de difusão da informação[...]” (GUINCHAT; MENOU, 1994, p.133). Elas estabelecem relações hierárquicas e remissivas que auxiliam tanto o indexador, nas operações de entrada do documento, quanto o usuário na identificação de conceitos relacionados, nas operações de busca e disseminação da informação.

Contudo, os custos elevados na elaboração e atualização das LDs, a necessidade de treinamento regular dos indexadores para uso das LDs, o vocabulário rígido das LDs e a desatualização devido aos avanços das áreas cobertas pelas LDs podem se apresentar como desvantagens desse tipo de linguagem, como salienta Guinchat e Menou (1994, p. 153):

A elaboração de uma linguagem documental fundamenta-se inicialmente na análise detalhada do sistema de informação. É necessário construir um Instrumento para uma situação bem precisa, e não apenas escolher e reunir palavras logicamente satisfatórias. Feita a análise do sistema, examinadas as linguagens existentes e tomada a decisão de adaptar ou de construir uma linguagem, deve-se fazer um plano de trabalho em função dos meios humanos, materiais e financeiros existentes. A adaptação ou a construção de uma linguagem é uma tarefa que leva vários meses ou alguns anos e necessita da colaboração de várias pessoas.

Contudo, essa Linguagem documentaria ainda é a melhor opção para uma descrição precisa dos documentos. Para que ocorra essa descrição, as regras e uso da Língua Portuguesa devem ser apreendidas e respeitadas pelos profissionais da informação atuantes em unidades de informação que utilizam a Língua Portuguesa na Indexação de seus documentos. O primeiro aspecto gramatical a ser observado é a criação dos morfemas das palavras, ou seja, a Morfologia.

4.2.3. Morfologia

Consiste na linha de pesquisa que estuda os aspectos morfológicos de condições de ocorrência das palavras, “[...] o estudo da estrutura dos morfemas, ou seja, dos constituintes das palavras. A morfologização é o processo de criação e alteração dessas estruturas [...]” (CASTILHO, 2010, p. 144). Exemplo:

(1) Julia completou o álbum.

No exemplo, podemos observar que a palavra ‘*completou*’ é formada por mais de um elemento: a sequência *complet-* mais a sequência *-ou*. A primeira pode gerar outros elementos lexicais (substantivos, verbo, advérbio) como *completo*, *completamente*, *completada* etc.; e a segunda, *-ou*, ocorre em outros verbos como *amou*, *desmanchou* etc. Ressalta-se que esta formação de palavras não é aleatória, “Existem também regras que governam a associação dessas partes de palavras (denominadas **morfemas**), e que impedem a formação de palavras como *⁵*desprezi*, **Ricardou* ou **Ricardível*.” (PERINI, 2009, p. 49-50).

Segundo Perini (2009), a Morfologia também estuda a associação de morfemas, essas associações de morfemas podem ser diversas:

⁵ Na Linguística convencionou-se que para construções agramaticais, não atestadas ou inexistentes na língua são indicadas por asterisco.

a) Quanto ao grau, no aumentativo e diminutivo, por exemplo:

(2) Esta casa é linda.

(3) Este casarão é lindo.

b) Quanto à flexão temporal de verbos, por exemplo:

(4) Ontem choveu em São Paulo.

(5) Hoje chove no Rio de Janeiro.

c) Quanto ao gênero de substantivos, por exemplo:

(6) Félix é o gato da Maria.

(7) Luna é a gata da Luiza.

d) Quanto à flexão de número, por exemplo:

(8) a folha da mangueira molhou com a chuva.

(9) as folhas da mangueira molharam com a chuva.

Por meio da Morfologia, entendemos os processos de formação de palavras, respeitando a ordem de junção de morfemas específicas da língua. De forma semelhante, pode-se ver que para formar frases ou orações existe uma ordem de distribuição das palavras, uma em relação a outra, a qual é expressa nos aspectos sintáticos da linguagem.

4.2.4. Sintaxe

A Sintaxe caracteriza-se pelo estudo da organização e das relações entre sintagmas (palavras) em uma frase. No português, as frases geralmente seguem uma ordem que é sujeito-verbo-adjetivo, por exemplo:

(10) Eu sou brasileira.

Porém, em outras línguas, como o coreano, a ordem da frase é sujeito-adjetivo-verbo, por exemplo:

(11) 저는 브라질 사람이에요. (Eu brasileira sou).

Levando em consideração o exemplo (10), pode-se observar a maneira como se associam as palavras para formar a frase. Assim, existe uma regra pela qual a terminação de *sou*, de certo modo, depende do elemento que se coloca no lugar de *Eu*; tanto é assim que, se em lugar de *Eu*, colocarmos *nós*, *sou* terá de se transformar em *somos*. Pode-se, além disso, notar que na frase (10) o elemento que governa a forma de *sou* ocorre em primeiro lugar na frase, e que modificações no último elemento da frase (*brasileira*) não afetam a forma de *sou*. Uma terceira observação é que existe um pequeno número de palavras que só podem ocorrer no lugar de *Eu*, e outras, em número igual, que só podem ocorrer no lugar de *brasileira*. *Ele*, *nós* só ocorrem no lugar de *Eu*, e *inglesa*, *bonita* só no lugar de *brasileira*. Todas essas observações têm a ver com a estruturação interna da frase, e constituem o estudo denominado sintaxe.

Como visto acima, essa organização da sentença tem regras, e Castilho (2010) ressalta a importância de haver limites nessa organização para a criação das estruturas nas sentenças. Com as palavras respeitando as regras de formação de palavras da língua e arranjadas de forma prevista pelo sistema dessa língua em uma frase, precisa-se verificar se essa frase faz sentido na língua portuguesa. O aspecto gramatical responsável pelo estudo sistemático do significado é a Semântica.

4.2.5. Semântica

Constitui-se, de acordo com Cançado (2008), o estudo do significado das línguas, ou seja, o sentido ou a significação dos signos linguísticos. Vejamos o exemplo (10): ao estudar uma frase como (10) *Eu sou brasileira*, pode-se levar em conta o significado transmitido por (10). Por exemplo, pode-se observar que *Eu* provavelmente designa uma pessoa/ ser humano, e *Brasileira* uma nacionalidade; que a pessoa é da nacionalidade *brasileira*, e não *Portuguesa*; que o fato de a pessoa ser brasileira acontece no presente, e assim por diante. Traços de significado como esses são, em parte, o resultado da aplicação de certas regras, que integram a Semântica.

Na análise Semântica, existem semas (unidades mínimas de significado) nucleares, que contêm a informação principal da frase, e acessórias/satélites, que ampliam o sentido da frase, porém não carregam a ideia principal.

Tomemos, como exemplo, a sentença "João chegou cedo". "João" desempenha a função sintática de sujeito, a função semântica de agente e a função pragmática de tema. Primeiro as funções semânticas são associadas com os predicados no léxico (por exemplo, agente + "chegar"), e o núcleo de uma sentença (no nosso exemplo, "João chegou") pode ser ampliado por satélites ("cedo", nesse caso) (MARTELOT, 2013, p. 162).

É nesse campo do conhecimento que se estudam algumas noções passíveis de dificuldades para indexadores como a Sinonímia, Antonímia, Homonímia e Polissemia. A Sinonímia é gerada quando dois signos linguísticos possuem o mesmo, ou muito próximo, significado. Por exemplo, nas frases em que os signos 'garoto' e 'menino' designam um jovem do sexo masculino:

(12) Aquele garoto canta muito bem.

(13) Aquele menino canta muito bem.

A Antonímia é gerada quando apresenta dois signos relacionados semanticamente, porém com significados/ideias opostos ou contraditórias. Como exemplificado nas sentenças em que a palavra 'quadrada' significa que o objeto tem a forma de um quadrado e 'redonda' significa que o objeto tem a forma circular ou oval, que é oposta à 'quadrada':

(14) Esta figura geométrica é redonda.

(15) Essa figura geométrica é quadrada.

A segunda dicotomia é a Homonímia e Polissemia. A primeira ocorre quando dois signos que possuem significados diferentes são escritos ou pronunciados da mesma forma. Por exemplo, o signo 'banco', que possui tanto o significado de instituição financeira, como de lugar em que se senta. Em ambos os significados, o termo é escrito e pronunciado de forma semelhante. A segunda ocorre quando o signo possui mais de um significado/sentido e possui a mesma grafia, como por exemplo, a palavra 'rede', que pode ser rede de deitar, rede elétrica, rede de computadores, etc. Em alguns momentos, Homonímia e Polissemia podem se confundir, porém em alguns casos a diferença é bem clara:

Essa diferença sempre deu problemas. Mas os casos mais extremos pareciam claros: a forma fonológica e gráfica *manga* pode designar uma fruta, ou então uma parte da roupa. Aí se via um caso evidente de homonímia: teríamos duas palavras, ou seja, dois itens lexicais, que se pronunciam e escrevem da mesma maneira. Já os vários significados da palavra *centro* (lugar geométrico; parte da cidade; local onde se desenvolve uma atividade, como *centro comercial*; posição política etc.) se vinculariam a um único item lexical, com diversas acepções: um caso de polissemia (PERINI, 2016, p. 542).

Quando falamos de Semântica, além do significado temos a hierarquização, como nas classes de espécie-gênero da biologia, e estruturação do campo semântico, ou seja, signos linguísticos com significados interligados. Eles são expressos por redes de palavras relacionadas e podem ser observados, por exemplo, na distinção de Tema e Rema:

Por *tema* de um enunciado se entende a parte que se refere ao que já é conhecido ou dado no contexto (também chamado às vezes, por outros teóricos, de *tópico* ou *assunto psicológico*). Por *rema*, a parte que veicula informação nova (o *comentário* ou *predicado psicológico*). (WEEDWOOD, 2002, p. 142)

Essas redes compõem campos semânticos, de acordo com Cintra, que ocorrem através do “[...]agrupamento das palavras em torno de um nó de significação, de um campo de semas.” (CINTRA, 1983, p. 12). Esses nós podem ser por relações hierárquicas (genérica e partitiva) e associativas como definidas pela norma ISO 704 (2000, p. 5 -10):

Uma relação genérica ocorre entre dois conceitos quando a intensão do conceito subordinado (conceito específico) inclui a intensão do conceito superordenado (conceito genérico) mais uma característica adicional (ISO 704, 2000, p. 5).

Uma relação partitiva ocorre quando o conceito superordenado representa um inteiro, enquanto o conceito subordinado representa parte deste inteiro. As partes, juntas, formam um inteiro. O conceito superordenado em uma relação partitiva é chamado o conceito compreensivo (*comprehensive concept*) e o conceito subordinado é chamado conceito partitivo. Conceitos subordinados em um mesmo nível, que compartilham a mesma dimensão são chamados conceitos coordenados (ISO 704, 2000, p. 9).

A relação partitiva e a relação genérica podem ser expressas em séries verticais e horizontais. As partes de um inteiro podem ser de natureza similar (átomo e molécula de oxigênio) ou diferentes entre si. Uma ou mais partes podem ser essenciais ou não-essenciais (opcionais). Para identificar as características essenciais dos conceitos partitivos é necessário determinar a intensão do conceito compreensivo (*comprehensive concept*) anteriormente. Segundo a norma, um sistema conceitual partitivo nem sempre permite uma análise completa de um conceito. Se o conceito partitivo não é particular ao conceito compreensivo, então a extensão de um conceito partitivo não explica completamente e as características de sua intensão podem se perder. Um conceito partitivo pode ser definido com base de uma relação partitiva apenas se a extensão completa e as características da intensão puderem ser determinadas (ISO 704, 2000, p. 10).

Já as relações associativas são relações não hierárquicas. Existe uma relação associativa quando uma conexão temática pode ser estabelecida entre conceitos em virtude da experiência. Algumas relações associativas ocorrem quando a dependência é estabelecida

entre conceitos com respeito a sua proximidade no espaço ou no tempo. Algumas relações envolvem eventos no tempo, como a dependência de um processo no tempo ou sequência; outras relações envolvem causa e efeito (ISO 704, 2000, p. 10). (CAMPOS, 2009, p. 63 - 64)

Outro fenômeno observado é a Ambiguidade, e é muito comum ocorrer na linguagem natural. O sentido dos signos em certas frases pode ser confuso ao ouvinte ou leitor, pois uma mesma sentença pode ter duas interpretações. Vejamos a sentença a seguir:

(16) Ana encontrou o gerente da loja com o seu irmão.

Essa sentença pode ser interpretada como ‘Ana’ encontrou o ‘gerente da loja’ com o irmão dele, como também que ‘Ana’ com o irmão dela encontraram o ‘gerente da loja’. Nesses casos, o contexto em que a sentença foi proferida irá ser determinante para a interpretação exata da informação:

Entender o que o professor falou em cada contexto específico parece envolver dois tipos de conhecimento. Por um lado, devemos entender o que o professor falou explicitamente, o que a sentença em português *a porta está aberta* significa; a esse tipo de conhecimento, chamamos de semântica. A semântica pode ser pensada como a explicação de aspectos da interpretação que dependem exclusivamente do sistema da língua e não, de como as pessoas a colocam em uso; em outros termos, podemos dizer que a semântica lida com a interpretação das expressões linguísticas, com o que permanece constante quando uma certa expressão é proferida. Por outro lado, não conseguiríamos entender o que o professor falou, se não entendêssemos também qual era a intenção dele ao falar aquela expressão para determinada pessoa em determinado contexto; a esse tipo de conhecimento, chamamos de pragmática. A pragmática estuda os usos situados da língua e lida com certos tipos de efeitos internacionais. (CANÇADO, 2008, p.17)

Existem dois níveis de interpretação de uma sentença: o semântico, visto neste tópico, que se refere à interpretação da palavra em si e dentro da frase, e o pragmático, referente ao significado em um determinado contexto, que será visto a seguir.

4.2.6. Pragmática

De acordo com Cançado (2008), a Pragmática é a disciplina que estuda a linguagem além de seus aspectos morfossintáticos e semânticos, observando a sua utilização em determinada situação comunicativa. Ela tem a incumbência de inferir, a partir do contexto, as informações implícitas, não expressas, na comunicação verbal, e, segundo Perini (2009), ocupa-se de entender as relações dos enunciados linguísticos com o contexto extralinguístico em que se inserem.

As inferências podem ser, por exemplo, anáforas em que “[...]o ouvinte infere que *isso* só pode estar se referindo a todas ocorrências ditas anteriormente[...]” (CANÇADO, 2008, p.131). Pode se observar na seguinte frase que o *isso* se refere às três frases/ acontecimentos anteriores:

(17) O voo foi tranquilo, rapidamente conseguiu um *uber*, o hotel era aconchegante. Acho que isso realmente fez a viagem ser maravilhosa para a Estela.

A inferência vai além da comunicação verbal. A comunicação não-verbal pode expressar também sentido em determinadas situações. Ao analisar a frase:

(18) Eu não posso nem falar em *kpop*.

O ouvinte pode inferir que o emissor, ou gosta muito de kpop que só de falar quer escutar e assistir aos vídeos, ou que o emissor não gosta nem um pouco de kpop que nem quer ouvir falar. Nestes casos, Cançado (2008) salienta que o ouvinte terá que recorrer a interpretação de gestos, entonações, expressões faciais, etc.

O contexto linguístico influencia a forma pronominal a ser usada, a forma morfológica de verbos e substantivos, como será expressa sintaticamente uma frase para a completa recepção pelo leitor, como expressar para que seja entendido sem ambiguidade pelo receptor da mensagem. Porém, os erros pragmáticos não infringem as regras da Fonologia, Sintaxe, Morfologia ou Semântica, pois ela não faz parte da estrutura da língua, mas ela leva em conta as intenções do falante/escritor e a recepção da mensagem pelo ouvinte ou leitor.

Portanto, observados os aspectos gramaticais acima, pode-se realizar a leitura e resumos de documentos respeitando a Morfologia, Sintaxe, Semântica e interpretando o contexto através da Pragmática, porém para representar ideias e conceitos dos documentos em termos é necessário mais um aspecto da Linguística, a saber, a Terminologia.

4.2.7. Terminologia

A Terminologia é entendida como um campo interdisciplinar que tem como objetivo estudar a denominação de conceitos, ou seja, termos, por meio de seus aspectos teóricos e metodológicos, segundo Medeiros (1986). Sendo *termo* e *conceito* definidos como:

3.1 – CONCEITO. [...]Os conceitos são, portanto, definidos como unidades do conhecimento, identificadas através de enunciados verdadeiros sobre um item de referência e representados por uma forma verbal.

3.2 – TERMO. Denomina-se termo um signo lingüístico que representa um conceito identificado na estrutura conceitual de um campo específico do conhecimento. Distinguem-se os termos das palavras da linguagem comum pela relação de univocidade (a um conceito corresponde apenas uma denominação) e de monorreferencialidade (a uma denominação corresponde apenas um conceito), que são estabelecidas entre conceito e termo. (MEDEIROS, 1986, p. 2)

O *termo* e *conceito* abordados na Terminologia estão intrinsecamente ligados à dicotomia saussureana:

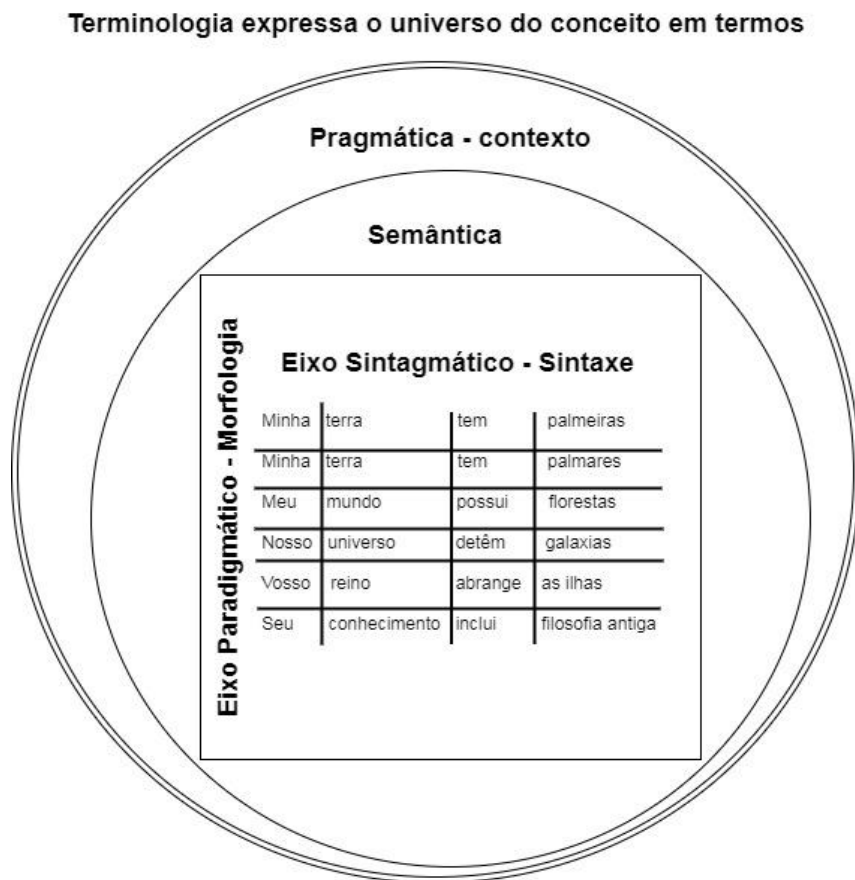
A sequência termo à denominação + conceito é nada mais que outra fórmula saussureana do signo metamorfoseado de seu enunciado clássico signo à significante + significado em um novo e mesmo esquema básico percebido meta terminologicamente. Em sua essência, a equação do mestre genebrino, que está na origem de toda a lingüística do século XX, resta imutável. Ela foi simplesmente adaptada à ecologia das LSP.

LG: signo significante + significado

LSP: termo denominação + conceito (BOULANGER, 1995, p. 3)

A Terminologia se faz necessária para representar o conteúdo dos documentos e utiliza-se dos conceitos vistos acima: Linguagem natural, Linguagem documentária, Morfologia, Sintaxe, Semântica e Pragmática para criar Termos que representem os conteúdos dos documentos. O emprego de um *termo* adequado e significativo de uma informação depende do conhecimento que o profissional tem da gramática e o contexto da língua, como se observa na seguinte ilustração:

Figura 1 - Universo Linguístico.



Fonte: Imagem feita pelo autor.

Por fim, a Terminologia incube-se de técnicas para usar morfemas, estruturados sintaticamente, expressando um sentido em si e em um determinado contexto do conhecimento, a fim de expressar conceitos em termos para posterior recuperação dessa informação. Para o uso adequado da Terminologia, quatro tipos de garantias devem ser observadas:

1. A Garantia Literária, que se refere aos termos presentes no próprio documento “Uma vez que a literatura de uma área pode ser definida, as expressões e termos nela contidos são indicativos da temática e tornam-se candidatos à inclusão no vocabulário da linguagem” (SVENONIUS, 2000, apud MOREIRA; MOURA, 2006, p. 6). Vale ressaltar que só essa garantia não é suficiente, pois um termo usado pelo autor do texto pode não corresponder aos termos utilizados na etapa de busca de informações.

2. A Garantia do Usuário, que está relacionada aos termos utilizados pelos usuários para recuperar a informação, “É importante acatar o vocabulário dos usuários e através dele conduzir as requisições que eles fazem aos descritores de um vocabulário mais especializado. Mas, sabe-se também que muitas vezes alguns usuários se perdem em palavras no instante da busca.” (MOREIRA; MOURA, 2006, p. 6), pois são usuários diversos, sendo necessário incluir todas as denominações pelo qual o conceito é conhecido.

3. A Garantia de Uso, referente à linguagem empregada pelo leitor, que é tanto usuário como criador de literatura. Esse leitor “[...]deveria ser empregado como uma fonte para se conhecer a linguagem mais apropriada para a indexação[...].” (MOREIRA; MOURA, 2006, p. 6), por ela ser mais próxima da realidade tanto do usuário como a do documento.

4. A Garantia Estrutural refere-se à linguagem estruturada de Tesouros, “São termos que facilitam elos em uma hierarquia de termos ou colaboram para que seja possível dispor um conjunto mais específico de termos. [...], ou seja, termos cuja colocação encontra justificativa na estrutura do tesouro” (MOREIRA ; MOURA, 2006, p. 7), garantindo o encontro de Termos dentro de campos semânticos e recuperação de itens relacionados.

Após o conhecimento dos métodos e conceitos linguísticos descritos acima, descreveremos com mais detalhes as etapas de Indexação e relacioná-las-emos com os tópicos abordados acima no próximo capítulo, dando ênfase nos momentos em que se fazem necessários os conhecimentos da Linguística.

5. ETAPAS DA INDEXAÇÃO E LINGUÍSTICA

Neste trabalho adota a concepção de Indexação segundo Lancaster (2004, p. 8-9), que “envolve duas etapas principais: 1. Análise conceitual, e 2. Tradução”. Essas etapas serão descritas a seguir e relacionadas com os conceitos da Linguística vistos acima.

5.1. ANÁLISE CONCEITUAL.

Na etapa de Análise conceitual, objetiva-se examinar a informação do documento, a fim de estabelecer o conceito de seu conteúdo através de uma leitura documentária. Fujita (2003) salienta que essa leitura documentária ocorre em três estágios: a Compreensão do texto como um todo, a Identificação e a Seleção de conceitos para a indexação.

No estágio da Compreensão do texto como um todo, é realizada uma leitura documentária dando atenção, segundo Fujita (2003), para “[...]o título, introdução e as primeiras frases dos capítulos e parágrafos; ilustrações, tabelas, diagrama e suas explicações; conclusão; palavras ou grupos de palavras sublinhadas ou impressas com tipo diferente”.

Essa leitura documentária é muitas vezes realizada com rapidez devido à velocidade que o processamento técnico deve ter no dia a dia da atividade. Por esta razão, Fujita (2004) ressalta que a leitura documentária, diferentemente da leitura comum, exige técnicas para sua execução, sendo importante a capacitação profissional, conhecimento e experiência para a elaboração de produtos documentários. Dentro desse conhecimento para a elaboração de produtos documentários, destacam-se os aspectos linguísticos abordados no capítulo 4.

Um dos aspectos é o conhecimento semântico utilizado para a compreensão do significado dos símbolos utilizados (texto, ilustração, gráfico, etc.), e o significado que o símbolo possui na frase. Esse conhecimento semântico ajuda a resolver dificuldades de compreensão textual como a *homonímia* e *polissemia*, muito presentes na Linguagem natural. Como no caso da palavra *manga* que pode designar uma fruta, ou então uma parte da roupa, o que fará compreender qual o sentido que está sendo utilizado no texto é a interpretação semântica deste.

Outra dificuldade de compreensão é a *ambiguidade*. O sentido dos signos em certas frases pode ser confuso, uma mesma sentença pode ter duas interpretações, como no exemplo: Ana encontrou o gerente da loja com o seu irmão. Essa sentença pode ser interpretada como ‘Ana’ encontrou o ‘gerente’ que estava acompanhado do irmão dele, como também que ‘Ana’ junto com

o irmão encontraram o ‘gerente’. Nesse exemplo, o estudo da Pragmática se faz necessário, pois ela tem a incumbência de inferir as informações implícitas, não expressas, na comunicação verbal.

Segundo Perini (2009), a Pragmática ocupa-se de entender as relações dos enunciados linguísticos com o contexto extralinguístico em que se inserem, ou seja, o tom de voz utilizado, quem proferiu a mensagem, em qual local. O estudo dessas informações auxiliará o indexador a interpretar a informação mais precisamente possível.

No estágio de *identificação de conceitos*, o indexador levanta conceitos de forma lógica concomitante com o seu conhecimento e prática numa determinada área, e através da frequência com que termos aparecem no documento. Esses termos que aparecem frequentemente compõem o que Kato (1982, apud Cintra 1983, p.7) chama de memória rasa: “[...]dá conta de algo já introduzido no texto e que mantém, dentro dele, o “status” de informação velha”.

Os termos que aparecem mais no documento constituem o foco deste, e os termos menos frequentes são complementares. Essa percepção da frequência de termos em um documento, relevante para o ato de Identificação de conceitos, é abordada no campo da Semântica como *Tema e Rema*:

Por *tema* de um enunciado se entende a parte que se refere ao que já é conhecido ou dado no contexto (também chamado às vezes, por outros teóricos, de *tópico* ou *assunto psicológico*). Por *rema*, a parte que veicula informação nova (o *comentário* ou *predicado psicológico*). (WEEDWOOD, 2002, p. 142)

Ao levantar os conceitos do documento, o indexador pode identificar pontos principais por meio de perguntas, as quais de acordo com Tálamo (1987 apud FUJITA, 2003, p. 77) são: “Quem? (Ser), O que? (Tema), Como? (Modo), Onde? (Lugar), e Quando? (Tempo)”. Para poder responder essas perguntas o indexador precisa fazer uma análise morfológica, sintática, pragmática e semântica das informações contidas nos documentos.

A Morfologia auxilia a identificar os signos e sua classe gramatical, como um adjetivo, um substantivo, um verbo, um artigo, um pronome, um advérbio, preposição, conjunção, numeral e interjeição, que ajudam a identificar as respostas às perguntas: “*onde?*”, “*quando?*”, “*quem?*”. A pergunta “*onde?*” é respondida com um advérbio de lugar, a pergunta “*quando?*” será respondida com um advérbio de tempo, e a pergunta “*quem?*” geralmente será um substantivo ou pronome e não um numeral, adjetivo ou interjeição, por exemplo.

A Sintaxe é relevante para identificar as funções desempenhadas pelas palavras nas orações e as relações sintagmáticas estabelecidas entre os signos no texto, por exemplo, a função de sujeito, predicado, predicativo do sujeito, predicativo do objeto, complemento nominal, agente da passiva, objeto direto, objeto indireto, adjunto adverbial, adjunto adnominal e aposto, que ajudam a identificar as respostas às perguntas: “*onde?*”, “*quando?*”, “*quem?*”. A pergunta “*onde?*” é respondida com o adjunto adverbial de lugar, a pergunta “*quando?*” será respondida com um adjunto adverbial de tempo, e a pergunta “*quem?*” geralmente será o sujeito ou objeto da frase e não um aposto ou predicativo do objeto por exemplo.

Na teoria das infinitas frases possíveis dos falantes de Charles Fillmore (1968), as frases contêm sintagmas ligados entre si desempenhando funções sintáticas, porém com uma função semântica (categorias) que varia de acordo com o caso. Essa função semântica nos ajuda a identificar o “*quem?*”.

Um pronome pode exercer a função sintática de sujeito, mas para entender se ele foi o sujeito que realizou a ação ou recebeu a ação, terá que realizar uma análise semântica. No exemplo: ‘Ela chutou a bola em João’, se for perguntado quem chutou a bola, a resposta será “*Ela*”, sendo o sujeito agente, mas se a pergunta for quem recebeu a bolada, a resposta será “*João*”, sendo este o sujeito afetado.

A Pragmática e a Semântica ajudam a entender o sentido ou a significação dos signos linguísticos no contexto intralinguístico e extralinguístico da informação, auxiliando a identificar as respostas às perguntas: “*o que?*”, “*como?*”. No contexto intralinguístico, com o auxílio da Semântica, interpretamos os símbolos linguísticos e as frases, e levantamos o “*o que?*” se está abordando, além de entendermos “*como?*” ocorre tal ação ou fenômeno no texto.

No contexto extralinguístico, para se responder a questão “*o que?*”, além de ser necessário interpretar o texto em si, precisa-se entender sobre qual ponto de vista foi escrito, pois os textos podem trazer frases ambíguas. Por isso a Pragmática é muito importante, pois o contexto da produção do documento pode modificar a sua interpretação; por exemplo, na frase: ‘Está chovendo canivete’, se for um conto de fadas ou filme essa frase pode realmente descrever que está chovendo canivete, porém se for um texto jornalístico, significa que o reporte usou uma expressão linguística para descrever que está chovendo muito forte.

No estágio de Seleção de conceitos, é preciso ter em vista o objetivo da informação indexada. Esse objetivo está de acordo com a política de Indexação da Instituição. A política de

Indexação da Instituição é que decide “[...]se o tema extraído do documento será o mais específico, ou se considerará um nível mais genérico, baseando-se no perfil do usuário que estabeleceu atender” (FUJITA, 2003, p. 69).

Para compreender a relação de termos específicos e termos genéricos, investigam-se na Semântica os termos que compõem os campos semânticos, que de acordo com Cintra (1983, p. 12), ocorrem através do “[...]agrupamento das palavras em torno de um nó de significação, de um campo de semas.”. Como no seguinte exemplo do Vocabulário controlado do Senado Federal (2017), em que se observa uma relação entre os termos, os quais foram agrupados em torno do assunto “cooperativa”:

Cooperativa
 UP Empresa cooperativa
 UP Sociedade cooperativa
 TE Cooperativa agrícola
 TE Cooperativa central
 TE Cooperativa de consumo
 TE Cooperativa de crédito
 TE Cooperativa de eletricidade
 TE Cooperativa de eletrificação rural
 TE Cooperativa de produção
 TE Cooperativa de trabalho
 TE Cooperativa econômica
 TE Cooperativa escolar
 TE Cooperativa habitacional
 TE Cooperativa industrial
 TE Cooperativa pesqueira
 TR Kibutz TR Sociedade de capital
 TR Sociedade simples
 Class. 334 (VCB, 2017, p. 155)

As relações entre os termos nos Tesouros ocorrem nas diversas formas como denotados na norma ISO 704 citado no capítulo 3. Essas relações são expressas por símbolos e abreviaturas denotadas nas Diretrizes para o estabelecimento de Tesouros Monolíngues (AUSTIN; DALE, 1993, p. 15), em que as relações hierárquicas são representadas por TG (Termo Genérico) e TE (Termo Específico), as relações equivalentes são representadas por UP (Usado para) e Use (o termo que se segue é o termo preterido), que são relações estabelecida entre um termo autorizado e um não-autorizado, as relações associativas representadas por TR (Termo Relacionado), além de haver notas explicativas representadas por NE para ligar um termo (significante) ao seu conceito (significado). No exemplo acima, o termo é Cooperativa (TG) é usado para Empresa cooperativa e Sociedade cooperativa. O termo Cooperativa tem como termos específicos Cooperativa agrícola, Cooperativa central, etc. E tem como termo relacionado sociedade simples.

O tema do documento decorre da análise conceitual do conteúdo deste. A Tematicidade do conteúdo de um documento tem pelo menos dois focos segundo Fujita (2003), o do ponto de vista do autor, chamado de Tematicidade Intrínseca, e do ponto de vista do leitor, chamado de Tematicidade Extrínseca.

Ao analisar a Tematicidade do documento, percebe-se a relação com as garantias estudadas na Terminologia, em que a Tematicidade Intrínseca está relacionada à Garantia Literária, que se refere aos termos presentes no próprio documento; e a Tematicidade Extrínseca está vinculada à Garantia do Usuário, que diz respeito aos termos utilizados pelos usuários para recuperar a informação.

A Garantia de Uso, que se refere à linguagem que é empregada pelo leitor que é tanto usuário como criador de literatura, está relacionada à Tematicidade Extrínseca. Ela também leva em consideração o contexto (competência pragmática) que está sendo utilizado, sendo assim, selecionam-se conceitos que sejam relevantes e de conhecimento do usuário, e da instituição em que o documento está sendo indexado.

5. 2. TRADUÇÃO EM LINGUAGEM DOCUMENTÁRIA

Na etapa de Tradução, Lancaster (2004, p. 18) afirma que “[...]envolve a conversão da análise conceitual de um documento num determinado conjunto de termos de indexação.” Esses

termos de Indexação são, geralmente, extraídos de instrumentos de Indexação. Esses instrumentos podem ser Tesouros, Listas de cabeçalhos de assunto, Sistema de classificação, etc.

Ao traduzir os termos da Linguagem natural dos documentos para os termos das LDs, podemos usar, por exemplo, os Tesouros que são uma Linguagem documentária composta por termos/descriptores organizados em uma estrutura semântica, de uma área ou mais do conhecimento, e que demonstra relações dos descriptores com os não descriptores. Esse instrumento tem a função de controlar os termos/descriptores para auxiliar na Tradução da Linguagem natural dos documentos para uma linguagem sistêmica mais rígida, visando sua posterior recuperação e disseminação.

De acordo com as Diretrizes para o estabelecimento de Tesouros Monolíngues (1993, p. 14), os Tesouros possuem termos preferidos que são “termos utilizados consistentemente para representar conceitos no processo de indexação; também conhecidos como descriptores”, e termos não-preferidos que são:

[...]sinônimos ou quase-sinônimos de termos que não são relacionados a documentos, mas que aparecem como pontos de entrada em um tesouro ou índice alfabético, enviando o usuário, por meio de instrução (por exemplo, USE e VEJA), ao termo apropriado; também conhecidos como não-descriptores (AUSTIN; DALE, 1993, p. 14).

A linguagem natural muitas vezes apresenta polissemia, sinonímia, homonímia, antonímia como vistas no capítulo 4. No caso de polissemia, as linguagens documentárias conseguem contornar este problema devido ao fato de vincular uma forma (Significante), representada no termo preterido, a um único sentido (Significado) expresso em notas explicativas. Quanto à sinonímia e antonímia, esses problemas são contornados com a vinculação dos campos semânticos através de relações entre os termos indicados pelo UP e Use.

Já a homonímia, quando ocorre de forma parcial como em coser e cozer, não representa uma barreira pois as linguagens documentárias apenas lidam com a língua escrita, assim dissipando a homonímia. Porém, quando a homonímia é total, é necessária uma diferenciação ou utilização de um sinônimo de maior recuperação, que é a indicação do termo preterido no tesouro e sua nota explicativa.

Ao utilizar os termos dos tesouros para a tradução dos conceitos levantados na etapa de Análise conceitual, promove-se a Garantia Estrutural da Terminologia utilizada na Indexação, além de garantir o encontro de termos dentro de campos semânticos e recuperação de itens relacionados. Em certos casos, o indexador poderá ter que criar termos, como na criação de um Tesouro, para

isso é importante que este tenha domínio das condições de ocorrência e da formação dos signos, ou seja, um bom conhecimento morfológico da língua em uso.

Além de ter um bom conhecimento semântico, para vincular a palavra a um único significado, e pragmático para escolher o termo preterido que condiz com o uso e a literatura da área indexada, e com o auxílio da terminologia, os termos podem ser extraídos da própria linguagem natural, ou da literatura que se está indexando, promovendo, assim, a Garantia Literária de uso e do usuário.

Um dos produtos da Tradução são as palavras-chave que acompanham, geralmente, outro produto, o resumo. Segundo a NBR 6028, as palavras-chave são palavras representativas do conteúdo do documento, sendo preferentemente de vocabulário controlado. O resumo consiste em uma apresentação concisa dos pontos relevantes de um documento, devendo ressaltar o objetivo, o método, os resultados e as conclusões do documento, devendo ser composto de uma sequência de frases concisas.

Na produção das palavras-chave, as metodologias da Morfologia, Semântica se fazem necessária tanto para expressar esses termos no código linguístico como para ser relevante para o usuário que procurar por estes termos no sistema. Ou seja, os termos devem ser escritos sem erros gramaticais para que o usuário consiga recuperar, e os termos devem representar conceitos e ideias nos campos semânticos que o usuário possa vir a procurar.

Além disso, o indexador vai escolher, de acordo com a política da instituição, se as palavras-chaves serão empregadas de forma exaustiva abrangendo suficientemente o conteúdo de modo bastante completo, ou será mais específica, abrangendo apenas o conceito principal do documento, como denotado por Lancaster (2004).

Para isso os campos semânticos nos auxiliam a entender a relação entre os termos, quais são os mais genéricos e quais são os mais específicos através da análise dos significados destes, a fim de auxiliar o indexador na escolha entre um termo e outro. É importante observar a política da instituição a fim de manter a coerência intraindexador e entre indexadores.

Na produção do resumo, faz-se necessário conhecimento morfológico para flexão e formação correta dos signos linguísticos, sintático para promover coerência e coesão ao texto, semântico para deixar claro qual é o objetivo, o método, os resultados e as conclusões do

documento, e conhecimentos pragmáticos para ser um texto sem ambiguidades, na linguagem adequada para o usuário destinatário da informação.

A fim de garantir a recuperação da informação, utilizamos, de forma auxiliar, os conceitos da linguística para processar essa informação e entender o seu significado através da Semântica, e o seu contexto através da Pragmática. Além de traduzir a informação em termos através da Terminologia, respeitando aspectos morfossintáticos da linguagem a fim de que a informação seja disseminada. O uso dessa metodologia será exemplificado no capítulo seguinte.

6. PRÁTICA DE INDEXAÇÃO

Neste capítulo, foi realizada a indexação do texto jornalístico retirado do Jornal Correio Braziliense ⁶, intitulado “Polícia entra em confronto com manifestantes contra a Copa do Mundo”. A partir de uma leitura integral do texto, foi feita uma análise semântica para determinar o tema e o rema dos conceitos abordados no texto, conforme frequência em que eles apareciam no texto. O tema foi identificado como: Confronto no protesto entre manifestantes contra a Copa do Mundo e policiais ao redor do Estádio Mané Garrincha.

De forma semelhante, foi feita uma análise pragmática para compreender o texto em sua totalidade e poder resolver problemas de ambiguidade observados na análise semântica. Em seguida, foi realizada uma análise de cada parte do texto observando aspectos morfosintáticos, semânticos e pragmáticos para auxiliar a responder as 5 questões apresentadas por Tálamo (1987 apud FUJITA, 2003).

Essa indexação foi feita utilizando os conceitos da Linguística para auxiliar as etapas de Análise conceitual e Tradução em linguagens documentárias, para demonstrar as contribuições observadas nesse estudo. A seguir apresenta-se o texto escolhido de forma completa e fragmentado na coluna à esquerda, e a Análise conceitual com os conceitos linguísticos destacados na coluna da direita.

Quadro 1: Indexação de Texto Jornalístico a luz da Linguística.

Texto Jornalístico fragmentado	Análise conceitual
Título: Polícia entra em confronto com manifestantes contra a Copa do Mundo	Neste caso a ambiguidade permeia na dúvida de quem estava contra a Copa do Mundo: Os policiais ou os manifestantes? Duas interpretações podem ser inferidas: 1. A polícia estava contra a Copa do Mundo.

⁶ Publicado em: 27/05/2014, disponível em: https://www.correiobraziliense.com.br/app/noticia/cidades/2014/05/27/interna_cidadesdf,429630/policia-entra-em-confronto-com-manifestantes-contr-a-copa-do-mundo.shtml. Encontra-se na Integra no Anexo 1.

	<p>2. Os manifestantes estavam contra a Copa do Mundo.</p> <p>Ao ler os trechos: “<i>Cerca de 2,5 mil pessoas se reuniram nos arredores do estádio Mané Garrincha em um protesto contra a Copa do Mundo</i>” e “<i>Um grupo de manifestantes jogou restos do material da obra das ciclovias contra os militares</i>”, depreende-se pelo contexto (pragmática) que eram os manifestantes que estavam contra a Copa do Mundo, e não os policiais.</p> <p>Analisando semanticamente temos dois sujeitos como “<i>quem</i>”, os agentes “<i>Polícia</i>” e “<i>os manifestantes</i>”, pois ambos estavam no confronto.</p>
<p>Cerca de 2,5 mil pessoas se reuniram nos arredores do estádio Mané Garrincha em um protesto contra a Copa do Mundo.</p>	<p>Neste trecho, responde-se às perguntas “<i>onde</i>” e “<i>quem</i>”. Com a ajuda da semântica identifica-se “<i>arredores do estádio Mané Garrincha</i>” e “<i>Mundo</i>” como lugares.</p> <p>Porém, sintaticamente, apenas o vocabulário “<i>arredores do estádio Mané Garrincha</i>” é caracterizado como adjunto adverbial de lugar por estar relacionado ao verbo “<i>reuniram</i>”, enquanto “<i>Mundo</i>” é caracterizados como adjunto adnominal por complementar o sentido do objeto “<i>Copa</i>”.</p> <p>Para responder à pergunta “<i>onde?</i>”, neste determinado contexto, apenas o vocábulo “<i>arredores do estádio Mané Garrincha</i>” pode ser a resposta, pois ele indica local onde a ação ocorreu.</p> <p>Para responder o “<i>quem?</i>” utiliza-se a Teoria das infinitas frases possíveis, de Charles Fillmore, para entender quem é que saiu dos arredores do Mané Garrincha e seguiu pela via N1. Neste caso, é o sujeito agente “<i>Cerca de 2,5 mil pessoas</i>”.</p> <p>Morfologicamente, entende-se que 2,5 mil é uma quantidade e não uma data ou hora. E semanticamente, entende-se que “<i>Cerca de 2,5 mil pessoas</i>” são os manifestantes relatados no título do texto, pois são eles que se reuniram.</p>

	<p>Com a análise morfológica, sintática, pragmática e a semântica do texto completo responde-se à pergunta “<i>o que</i>” da seguinte maneira: “confronto no protesto entre manifestantes contra a Copa do Mundo e policiais ao redor do Estádio Mané Garrincha”, sendo este o tema central do texto.</p> <p>O tema irá ser retomado nos trechos a seguir, sendo o mais frequente no texto como visto na análise morfológica do texto. Somente a palavra manifestante apareceu 5 vezes no texto, já as palavras Copa do Mundo, policiais e Estádio Mané Garrincha aparecem pelo menos duas vezes com suas formas variantes.</p>
<p>O grupo saiu da Rodoviária do Plano Piloto por volta das 16h30 desta terça-feira (27/5) e seguiu pela via N1, bloqueando todas as seis faixas, no sentido Palácio do Buriti.</p>	<p>Nesta parte, necessita-se inferir, através da pragmática, o sujeito que responde à pergunta “<i>quem</i>”, o “<i>o grupo</i>” que se refere às “<i>2,5 mil pessoas</i>” anaforicamente, pois remete ao sujeito da frase anterior.</p> <p>Para responder as perguntas “<i>quando?</i>” e “<i>onde?</i>”, com a ajuda da semântica, identifica-se <i>Rodoviária do Plano Piloto e via N1, Palácio do Buriti</i> como lugares.</p> <p>Porém, sintaticamente, apenas os vocabulários “<i>Rodoviária do Plano Piloto</i>” e “<i>via N1</i>” são caracterizados como adjunto adverbial de lugar por estarem relacionados aos verbos “<i>saiu</i>” e “<i>seguiu</i>”, enquanto “<i>Palácio do Buriti</i>” é caracterizado como adjunto adnominal por complementar o significado do objeto “<i>sentido</i>”.</p> <p>Para responder à pergunta “<i>onde?</i>”, apenas os vocábulos <i>Rodoviária do Plano Piloto e via N1</i> podem ser a resposta, pois eles indicam locais onde a ação ocorreu.</p> <p>A pergunta “<i>quando?</i>” será respondida de acordo com a sintaxe com um adjunto adverbial de tempo. E, de acordo com a morfologia, será respondida com um advérbio de tempo que pode ser identificado no trecho como: “<i>por volta das 16h30 desta terça-feira (27/5).</i>”.</p>

	<p>Os números do trecho denotam uma hora e data, sendo de fácil identificação para o indexador devido a sua forma morfológica. Analisando semanticamente, depreende-se que a frase destacada acima indica um período no tempo, e analisando sintaticamente entende-se que esse período de tempo está relacionado ao verbo “<i>saiu</i>” caracterizado como um advérbio de tempo.</p> <p>O trecho faz parte do tema na identificação de assuntos, pois semanticamente entende-se como os manifestantes começaram o protesto, respondendo assim à pergunta “<i>como</i>”.</p>
<p>Policiais militares do Batalhão do Choque interditaram o acesso dos manifestantes ao Estádio Nacional, já na altura da Torre de TV.</p>	<p>Neste caso, para responder o “<i>quem?</i>”, entende-se que o vocábulo “<i>Policiais militares do Batalhão do Choque</i>” é o sujeito sintaticamente, e semanticamente é o agente da ação de interditar.</p> <p>Já o vocábulo “<i>manifestantes</i>” não é um sujeito, ele é um adjunto adnominal e entende-se semanticamente que ele possui o acesso, mas não participa da ação do verbo interditar, não se caracterizando como um “<i>quem?</i>”.</p> <p>Para responder à pergunta “<i>onde?</i>”, apenas o vocábulo “<i>na altura da Torre de TV</i>” pode ser a resposta, pois ele indica o local onde a ação ocorreu, sendo caracterizado sintaticamente como um adjunto adverbial, porque ele se relaciona com o verbo interditaram. Enquanto Estádio Nacional é um adjunto adnominal, pois complementa semanticamente o objeto “acesso”, não sendo assim o local que a ação ocorreu.</p> <p>O trecho faz parte do tema na identificação de assuntos, pois semanticamente entende-se como o conflito entre os policiais e os manifestantes começou, respondendo assim à pergunta “<i>como</i>”.</p>
<p>Eles usaram bombas de gás para dispersar os manifestantes, que fazem</p>	<p>Nesta parte, necessita-se inferir, através da pragmática, os sujeitos que respondem à pergunta “<i>quem</i>”, o “<i>Eles</i>” que se</p>

<p>parte de movimentos como Comitê Popular contra a Copa do Mundo, o Juntos, e do Movimento dos Trabalhadores Sem Teto (MTST). Também estiveram na manifestação os índios que fizeram parte de protestos na Praça dos Três Poderes nesta manhã.</p>	<p>refere aos policiais anaforicamente, pois remete ao sujeito da frase anterior.</p> <p>Além disso, infere-se que “<i>os manifestantes</i>” engloba as pessoas que fazem parte “<i>de movimentos como Comitê Popular contra a Copa do Mundo, o Juntos, e do Movimento dos Trabalhadores Sem Teto (MTST)</i>” e os índios.</p> <p>Analisando sintaticamente temos dois sujeitos, o agente “<i>Eles</i>” e o paciente “<i>os manifestantes</i>” com todos que estão inferidos em seu conceito.</p> <p>O primeiro trecho faz parte do tema na identificação de assuntos, pois semanticamente entende-se como ocorreu o confronto entre os policiais e os manifestantes, respondendo assim à pergunta “<i>como</i>”.</p> <p>Para responder à pergunta “<i>onde?</i>”, o vocábulo “<i>Praça dos Três Poderes</i>” podem ser a resposta, pois indica o local onde fizeram protestos, sendo caracterizado sintaticamente como adjunto adverbial.</p> <p>A pergunta “<i>quando?</i>” será respondida, de acordo com a sintaxe, com um adjunto adverbial de tempo, que é identificado semanticamente no período de tempo “<i>nesta manhã</i>”.</p> <p>Porém, ao identificar os assuntos, o segundo trecho que responde às perguntas “<i>quando?</i>” e “<i>onde?</i>” é entendido como rema, pois veicula uma informação nova, e não como tema, uma vez que este seria a manifestação e não os protestos na Praça dos Três Poderes. Sendo assim, não se considera como resposta do onde e quando.</p>
<p>Um grupo de manifestantes jogou restos do material da obra das ciclovias contra os militares.</p>	<p>Analisando sintaticamente temos dois sujeitos como “<i>quem</i>”, que desempenham a função semântica de agente “<i>Um grupo de manifestantes</i>” e paciente “<i>os militares</i>”.</p> <p>O trecho faz parte do tema na identificação de assuntos, porque semanticamente entende-se como ocorreu o confronto</p>

	entre os policiais e os manifestantes, respondendo assim à pergunta “ <i>como</i> ”.
Um policial da cavalaria foi atingido por uma flecha na perna, lançada por índios.	<p>Sintaticamente temos dois sujeitos como “<i>quem</i>”, que desempenham a função de agente “<i>índios</i>” e paciente “<i>um policial</i>”.</p> <p>O trecho faz parte do tema na identificação de assuntos, pois semanticamente entende-se como ocorreu o confronto entre os policiais e os manifestantes, respondendo assim à pergunta “<i>como</i>”.</p>
Segundo a Polícia Civil, duas pessoas foram detidas e encaminhadas à Delegacia de Polícia Especializada (DPE), entre elas, um homem que jogava pedras contra um ônibus.	Sintaticamente, temos três sujeitos, que desempenham as funções semânticas de agentes “ <i>Polícia Civil</i> ” e “ <i>um homem</i> ” e paciente “ <i>duas pessoas</i> ”. Porém, ao identificar os assuntos, entende-se esse trecho como rema, pois, veicula uma informação nova, e não como tema, pois este seria a manifestação. Sendo assim, esses sujeitos não entrariam na resposta do “ <i>quem</i> ”.
Com o acesso ao Eixo Monumental bloqueado pelos manifestantes, alguns motoristas que já estavam na via tiveram que retornar de ré para não entrarem no protesto.	<p>Sintaticamente, temos dois sujeitos como “<i>quem</i>”, que desempenham a função semântica de agente “<i>manifestantes</i>” e paciente “<i>motoristas</i>”. Sintaticamente, denota-se o adjunto adverbial de lugar “<i>protesto</i>” relacionado ao verbo “<i>entraram</i>” como o “<i>onde</i>”.</p> <p>Porém, ao identificar os assuntos, compreende-se esse trecho como rema, pois, veicula uma informação nova, e não como tema, pois este seria a manifestação. Sendo assim, esses sujeitos e lugares do trecho não entrariam na resposta do “<i>quem</i>” ou “<i>onde</i>”.</p>
Alguns abandonaram os carros e fugiram pelo gramado central. O trânsito no Eixo Monumental provocou	Sintaticamente, temos dois sujeitos como “ <i>quem</i> ”, que desempenham a função semântica de agentes “ <i>alguns</i> ” e “ <i>o trânsito no Eixo</i> ”.

<p>reflexos nas pistas do Anexo dos Ministérios.</p>	<p>Sintaticamente, denota-se o adjunto adverbial de lugar “<i>nas pistas do Anexo dos Ministérios</i>” relacionado com o verbo “<i>provocou</i>” como o “<i>onde</i>”.</p> <p>Mas, ao identificar os assuntos, entende-se esse trecho como rema, pois, veicula uma informação nova, e não como tema, á que este seria a manifestação. Sendo assim, esses sujeitos e lugares do trecho não entrariam na resposta do “<i>quem</i>” ou “<i>onde</i>”.</p>
<p>O engarrafamento foi desde o Congresso Nacional até o cruzamento com a W3.</p>	<p>Sintaticamente temos um sujeito como o “<i>quem</i>”, que desempenha a função semântica de agente “<i>engarrafamento</i>”. Sintaticamente, denota-se os adjuntos adverbial de lugar “<i>Congresso Nacional</i>” e “<i>cruzamento com a W3</i>” relacionado com o verbo “<i>foi</i>” como o “<i>onde</i>”.</p> <p>Porém, ao identificar os assuntos, entende-se esse trecho como rema, pois, veicula uma informação nova, e não como tema, pois este seria a manifestação. Sendo assim, esses sujeitos e lugares do trecho não entrariam na resposta do “<i>quem</i>” ou “<i>onde</i>”.</p>
<p>O protesto começou de forma pacífica, com a participação de crianças segurando faixas e cartazes, mas depois o cenário foi mudando e algumas crianças firam vistas correndo, e chorando com seus pais e dentro de ônibus com a mão no rosto, na tentativa de não respirar os gases lançados durante o confronto.</p>	<p>Sintaticamente, temos três sujeitos como “<i>quem</i>”, que desempenham a função semântica de agentes “<i>protesto</i>”, “<i>crianças</i>”, “<i>seus pais</i>”.</p> <p>O trecho faz parte do tema na identificação de assuntos, pois semanticamente entende-se como ocorreu o desfecho do confronto entre os policiais e os manifestantes, respondendo assim à pergunta “<i>como</i>”.</p>

Fonte: Elaborado pela autora.

A seguir, relata-se a etapa de Tradução, que utilizou o Vocabulário Controlado do Senado Federal⁷ de 2017 devido a ser o vocabulário utilizado como base para várias instituições de informação. Este apresenta uma Linguagem documentária estruturada em campos semânticos, que padroniza a linguagem e evita duplicidade de significação, além de ligar termos a conceitos. Nesse ponto, a definição de Termo e Conceito, e Significado e Significante auxiliou para traduzir ideias e conceitos da Linguagem natural para a Linguagem documentária.

De forma semelhante, produziu-se um resumo observando a coerência e coesão da Sintaxe e Semântica, a clareza da Pragmática, e as garantias da Terminologia para com o leitor do texto. Foram trazidos os conceitos da Análise Conceitual levantados através das perguntas de acordo com Tálamo (1987 apud FUJITA, 2003) a fim de estruturá-lo em um parágrafo conciso e claro, como descrito na NBR 6028. Em ambos os casos, também se observou a forma de escrever os vocábulos na Língua Portuguesa, ou seja, a sua Morfologia.

1. Termos de indexação e Resumo à luz da Linguística.

Termos de indexação	
Termos e conceitos extraídos da Análise Conceitual	Termos traduzidos utilizando o VCB
Manifestantes;	Movimento de protesto;
Policiais;	Polícia;
Copa do Mundo;	Competição esportiva;
Confronto	Violência urbana.
Resumo	
Análise conceitual para redação do resumo	Resumo utilizando os conceitos linguísticos.
<p>O quê?</p> <p>Confronto no protesto entre manifestantes contra a Copa do Mundo e policiais ao redor do Estádio Mané Garrincha.</p> <p>Quem?</p>	<p>O confronto entre manifestantes contra a Copa do Mundo e policiais ocorreu ao redor do Estádio Mané Garrincha e começou por volta das 16h30 da terça-feira, dia 27 de maio de 2014. O grupo de manifestantes, formado por movimentos</p>

⁷ Disponível em:

https://www2.senado.leg.br/bdsf/bitstream/handle/id/532112/001106656_VCB_2017_08_20.pdf?sequence=1.

<p>“Cerca de 2,5 mil pessoas”, “Policiais militares do Batalhão do Choque”, “de movimentos como Comitê Popular contra a Copa do Mundo, o Juntos, e do Movimento dos Trabalhadores Sem Teto (MTST).”, “Um grupo de manifestantes”, “os militares”, “índios”, “um policial”, “protesto”, “crianças”, “seus pais”.</p> <p>Como?</p> <p>Começo da manifestação: “O grupo saiu da Rodoviária do Plano Piloto por volta das 16h30 desta terça-feira (27/5) e seguiu pela via N1, bloqueando todas as seis faixas, no sentido Palácio do Buriti.”</p> <p>Começo do conflito: “Policiais militares do Batalhão do Choque interditaram o acesso dos manifestantes ao Estádio Nacional, já na altura da Torre de TV.”</p> <p>Como ocorreu: “Eles usaram bombas de gás para dispersar os manifestantes, que fazem parte de movimentos como Comitê Popular contra a Copa do Mundo, o Juntos, e do Movimento dos Trabalhadores Sem Teto (MTST).”, “Um grupo de manifestantes jogou restos do material da obra das ciclovias contra os militares.”, “Um policial da cavalaria foi atingido por uma flecha na perna, lançada por índios.”</p> <p>Como terminou: “O protesto começou de forma pacífica, com a participação de crianças segurando faixas e cartazes, mas depois o</p>	<p>como Comitê Popular contra a Copa do Mundo, o Juntos, Movimento dos Trabalhadores Sem Teto e os índios, saiu da rodoviária do Plano piloto em direção ao Estádio, e perto da Torre de TV foram bloqueados por Policiais militares do Batalhão do Choque. O confronto ocorreu com o uso de bombas de gás pelos policiais contra os manifestantes, flechas e restos de materiais de construção atirados pelos manifestantes contra os policiais. A manifestação começou pacífica, mas terminou de forma violenta.</p>
--	--

<p>cenário foi mudando e algumas crianças foram vistas correndo, e chorando com seus pais e dentro de ônibus com a mão no rosto, na tentativa de não respirar os gases lançados durante o confronto.”</p> <p>Quando?</p> <p>“por volta das 16h30 desta terça-feira (27/5).”</p> <p>Onde?</p> <p>“arredores do estádio Mané Garrincha”, “Rodoviária do Plano Piloto” e “via N1”, “na altura da Torre de TV”.</p>	
---	--

Fonte: Elaborado pela autora.

A análise linguística contribuiu para contornar a ambiguidade apresentada logo no título, evitando, assim, a indexação errada do assunto. Também auxiliou a responder às cinco questões apresentadas por Tálamo (1987 apud FUJITA, 2003). O “quem?” é identificado pelo papel sintático de sujeito, o “quando?” é identificado sintática e morfologicamente por expressões de tempo que são adjuntos adverbiais, o “como?” é identificado semanticamente por expressões de modo e morfologicamente por gerúndios ou advérbios, o “o que?” é identificado semanticamente por meio da frequência com que as palavras aparecem no texto e como as outras ideias complementam o foco principal, e o “onde?” é identificado semanticamente, morfologicamente e sintaticamente como o ambiente em que a ação aconteceu, sendo este um advérbio ou substantivo relacionado ao verbo.

Essa análise contribui principalmente para a tradução, pois apenas entendendo o sentido dos termos das linguagens documentárias podemos nos aproximar dos conceitos levantados no texto, e assim traduzir de forma mais fidedigna. Por exemplo só com a leitura do texto entendemos que esse “confronto” ocorreu na cidade e podemos traduzir confronto para o termo da linguagem documentária “Violência Urbana”, não sendo este confronto de ideias ou posicionamentos.

Por fim, os conhecimentos linguísticos auxiliaram na produção do resumo, pois os conceitos levantados na análise conceitual, através das perguntas de Tálamo (1987 apud FUJITA, 2003), facilitaram a escrita de um parágrafo conciso e claro como descrito na NBR 6028. Os

conceitos linguísticos de coerência e coesão da Sintaxe e Semântica, a clareza da Pragmática, a forma de escrever da Morfologia, e as garantias da Terminologia para com o leitor do texto foram imprescindíveis e norteadores para a confecção do resumo, sendo de conhecimento necessário a indexadores.

7. CONCLUSÃO

No presente trabalho, procurou-se entender quais aspectos da Linguística auxiliam o indexador no processo de Indexação. Entende-se que como a Biblioteconomia usa dos signos linguísticos para interpretar documentos e representar a informação contida neles, surge a necessidade de compreender o significado dos signos, e como e quando usá-los de forma apropriada. Porém, o ingresso no curso de Biblioteconomia pode não ter conhecimentos prévios suficientes de Linguística para suprir essa necessidade. E essa lacuna de conhecimento pode prejudicar seu desempenho em tarefas que envolvem a representação e recuperação da informação.

Para atingir o objetivo de identificar os aspectos linguísticos necessários para a formação de indexadores, primeiramente, no capítulo “Indexação: aspectos conceituais”, abordou-se a evolução das correntes teóricas da Indexação na literatura especializada da área da Ciência da Informação, sendo relatado de forma sucinta essa evolução.

Discorreu-se sobre as definições de Termo e Conceito, que são importantes para a Terminologia da Linguística e para as etapas de Indexação, principalmente, no desenvolvimento e uso de Tesouros. Percebe-se que Termo e Conceito estão diretamente relacionados com a dicotomia de Saussure Significado e Significante, e que o ensino dessa dicotomia no curso de Biblioteconomia auxiliaria indexadores a compreendê-los melhor.

De forma semelhante, detalharam-se alguns parâmetros de qualidade da indexação que orientam a utilização de termos neste processo. Definiram-se as etapas de Indexação que foram utilizadas no presente trabalho, visto que a Indexação, de acordo com diferentes autores, pode ter diversas etapas. Por mais que tenham diversas formas de dividir essas etapas, pode-se simplificar em duas etapas: a de análise textual e de tradução, sendo essas as etapas escolhidas pela autora.

Seguidamente, em “Linguística: aspectos conceituais”, abordou-se a origem dessa ciência a fim de se entender o contexto por detrás de aspectos conceituais relevantes para a prática de Indexação de forma pontual. No mesmo capítulo, detalharam-se os aspectos linguísticos e exemplificou os seus conceitos a fim de esclarecer os pontos relevantes para os profissionais da informação e para as etapas de Indexação.

Os aspectos da Linguística foram encontrados em sua maioria em fontes bibliográficas não específicas da Ciência da Informação, devido algumas dessas fontes abordarem de forma superficial os aspectos, mesmo quando utilizam destes para auxiliarem em seus trabalhos. Em sua maioria, abordam apenas aspectos pragmáticos, terminológicos e semânticos, deixando de lado os aspectos sintáticos e morfossintáticos que são nucleares em análises linguísticas e, como demonstrado no último capítulo, determinante em alguns momentos para identificar o *Quem, O que, Como, Quando e Onde*.

Analisaram-se, em “Etapas da Indexação e a Linguística”, as duas etapas da Indexação, de acordo com a definição da UNISIST (1981) e de Lancaster (2004) que consistem em: Análise conceitual e Tradução, relacionando com os aspectos linguísticos. Essa relação foi feita de forma exemplificativa, dos pontos de maior convergência entre a Linguística e as etapas da Indexação, e não exaustiva, pois existem vários aspectos linguísticos a serem estudados mais detalhadamente que podem auxiliar o indexador leitor, como a Análise do discurso, Semiótica etc.

Considera-se que o objetivo de identificar quais aspectos da Linguística auxiliam o indexador no processo de Indexação sendo necessários à sua formação foi alcançado através das relações demonstradas no capítulo “Etapas da Indexação e a Linguística” entre os conceitos levantados nos aspectos conceituais das duas áreas. Com o alcance desse objetivo, pode-se demonstrar na prática, no capítulo “Prática de Indexação”, a indexação de um texto jornalístico a partir do enfoque linguístico, exemplificando como a linguística é fundamental para a identificar os conceitos e a traduzi-los de forma apropriada. Demonstrou-se, em forma de quadros, como a Morfologia, Sintaxe, Semântica, Pragmática e a Terminologia ajudaram na Indexação do texto.

Denota-se que os aspectos basilares da Morfologia, Sintaxe, Semântica, Pragmática e a Terminologia são necessários para a formação do indexador. Assim, é sugerido que se crie uma disciplina para abordar os aspectos linguísticos com foco para cientistas da informação, ou que em uma disciplina base de língua portuguesa junto com as disciplinas do curso de Biblioteconomia seja abordado de forma gradual os níveis de análise linguística. Em uma disciplina base de Língua Portuguesa seja abordado: noções de Linguística, Morfologia, Sintaxe. Em *análise da informação* seja abordado: Semântica, Pragmática. Em *Classificação*: Terminologia. Em *Indexação*: a leitura ativa, analítica e crítica de textos. Planejamento e produção de resumos, resenhas críticas e textos dissertativos - argumentativos.

Tendo em vista a representação da informação, Melo; Bracher (2011) salienta que o uso de formas linguísticas na representação da informação deve ser adaptado – a adaptação ocorre através da competência pragmática do indexador – ao momento presente, ao público, à natureza do documento escolhido, aos objetivos do produto intelectual em mãos para que ocorra uma comunicação eficaz. Para ser eficaz, a comunicação deve respeitar as regras do sistema linguístico comum entre o usuário e o indexador, gerando a recuperação da informação.

Portanto, o conhecimento linguístico é necessário a indexadores, pois auxilia o processo de leitura documental e levantamento de conceitos como a confecção de resumos. A informação deve ser analisada e interpretada de forma correta para poder ser representada de forma fidedigna, para isso o indexador deve ter um bom conhecimento da língua do documento. Por isso que se indica uma disciplina de Linguística voltada para indexadores a fim de proporcionar uma boa base de conhecimentos da norma culta e dos níveis de análise da linguística, visando auxiliar o processo de indexação e evitar erros.

8. REFERÊNCIAS

ASSIS, T. B. **O uso das linguagens natural e controlada na recuperação da informação na web: o caso das livrarias eletrônicas**. 2010. 90 f. Monografia (Bacharelado em Biblioteconomia) -Universidade de Brasília, Brasília, 2010. Disponível em: <<http://bdm.unb.br/handle/10483/1221>>. Acesso em 01 set. 2019.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS – ABNT. **NBR 12676**: métodos para análise de documentos: determinação de seus assuntos e seleção de termos de indexação: procedimento. Rio de Janeiro, 1992.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS – ABNT. **NBR 6028**: Informação e documentação - Resumo - Apresentação. Rio de Janeiro, 2003.

AUSTIN, D.; DALE, P. **Diretrizes para o estabelecimento e desenvolvimento de tesauros monolíngües**. Brasília: Ibict/SENAI, 1993. 86 p. Disponível em: <http://livroaberto.ibict.br/handle/1/731>. Acesso em: 11 nov. 2019.

BARANOW, U. G. Perspectivas na contribuição da lingüística e de áreas afins à ciência da informação. **Ciência da Informação**, [S.l.], v. 12, n. 1, june 1983. ISSN 1518-8353. Disponível em: <<http://revista.ibict.br/ciinf/article/view/191>>. Acesso em: 15 jul. 2019.

BOULANGER, J. C. Alguns componentes lingüísticos no ensino da terminologia. **Ciência da Informação**, [S.l.], v. 24, n. 3, dec. 1995. ISSN 1518-8353. Disponível em: <<http://revista.ibict.br/ciinf/article/view/570>>. Acesso em: 15 jul. 2019.

CAMPOS, J. A. G. **Análise conceitual sobre as relações semânticas em ciência da informação**: contribuições para o desenvolvimento de ontologias. 2009. 140 f. Dissertação (Mestrado em Ciência da Informação). Universidade Federal de Minas Gerais, Belo Horizonte, 2009. Disponível em: <http://hdl.handle.net/1843/ECID-7V4PLA>. Acesso em: 17 set. 2019.

CANÇADO, M. **Manual de semântica**: noções básicas e exercícios. 2. ed. rev. Belo Horizonte: Editora UFMG, 2008. 183 p. (Didática (Ed. UFMG)). ISBN 9788570416803.

CASTILHO, A. T. **Nova gramática do português brasileiro**. São Paulo, SP: Contexto, 2010. 768 p. ISBN 9788572444620.

CINTRA, A. M. M. Elementos de lingüística para estudos de indexação. **Ciência da Informação**, [S.l.], v. 12, n. 1, june 1983. ISSN 1518-8353. Disponível em: <<http://revista.ibict.br/ciinf/article/view/190>>. Acesso em: 15 jul. 2019.

FUJITA, M. S. L.. A representação documentária no processo de indexação com o modelo de leitura documentária para textos científicos e livros: uma abordagem cognitiva com protocolo verbal. **PontodeAcesso (UFBA)**, v. 17, p. 42-66, 2013. Disponível em: <<https://portalseer.ufba.br/index.php/revistaici/article/view/8135>>. Acesso em: 15 jul. 2019.

FUJITA, M. S. L. A identificação de conceitos no processo de análise de assunto para indexação. RDBCI: **Revista Digital De Biblioteconomia E Ciência Da Informação**, Campinas, v. 1, n. 1, p. 60-90, jul./dez. 2003 – ISSN: 1678-765X.

GUINCHAT, C.; MENOU, M. **Introdução geral às ciências e técnicas da informação e documentação**. 2. ed. corr. aum. Brasília: IBICT, 1994. 540 p.

INTERNATIONAL STANDARD ORGANIZATION. **ISO 704**: Terminology work: Principles and methods principles and methods of terminology. Geneva: ISO, 2000.

LANCASTER, F. W. **Indexação e resumos**: teoria e prática. 2. ed. Brasília: Briquet de Lemos, 2004.

LEIVA, I. G.; FUJITA, M. S. L. **Política de Indexação**. São Paulo : Cultura Acadêmica; Marília: Oficina Universitária, 2012. Disponível em: https://www.marilia.unesp.br/Home/Publicacoes/politica-de-indexacao_ebook.pdf. Acesso em: 11 nov. 2019.

MARTELOTTA, M. E. (Org.). **Manual de lingüística**. 2. ed. São Paulo, SP: Contexto, 2013. 254 p. ISBN 9788572443869.

MEDEIROS, M. B. B. Terminologia Brasileira de Ciência da Informação: uma análise. **Ciência da Informação**, Brasília, v.15, n. 2, p. 135-42, jul./dez. 1986. Disponível em:<<http://revista.ibict.br/index.php/ciinf/article/viewFile/1425/1046>>. Acesso em: 11 nov. 2019.

MELO, F. J. D.; BRÄSCHER, M. **Fundamentos da linguística para a formação do profissional de informação**. Brasília : Thesaurus, 2011. 123 p. ISBN 9788564494459.

MELO, M. A. F.; BRÄSCHER, M.. Termo, conceito e relações conceituais: um estudo das propostas de Dahlberg e Hjørland. **Ciência da Informação**, [S.l.], v. 43, n. 1, june 2015. ISSN 1518-8353. Disponível em: <<http://revista.ibict.br/ciinf/article/view/1419>>. Acesso em: 11 nov. 2019.

Ministério da Educação. **Português tem apenas 1,6% de aprendizagem adequada no Saeb**. Disponível em:<<http://portal.mec.gov.br/ultimas-noticias/389-ensino-medio-2092297298/68271- apenas-1-6-dos-estudantes-do-ensino-medio-tem-niveis-de-aprendizagem-adequados-em-portugues>>. Acesso em: 01 set. 2019.

MOREIRA, M. P.; MOURA, Maria Aparecida. Construindo tesouros a partir de tesouros existentes: a experiência do TCI – Tesouro em Ciência da Informação. **DataGramaZero - Revista de Ciência da Informação** - v.7 n.4 ago/06. Disponível em: <<http://www.brapci.inf.br/index.php/article/view/0000007757/27466a615ff97c8ea30886db7bc3b9f4>>. Acesso em: 11 nov. 2019.

PERINI, M. A. **Gramática descritiva do português**. 4.ed. São Paulo: Ática, 2009. (Básica universitária). ISBN 9788508055500. E-book. Acesso em: 30 ago. 2019.

PERINI, M. A. **Gramática descritiva do português brasileiro**. Petrópolis: Vozes, 2016.

PRODANOV, C. C.; FREITAS, E. C. **Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico**. 2. ed. Novo Hamburgo: Feevale, 2013. 276 p.

RAWLS, A. W. Os Estudos de etnometodologia de Garfinkel: uma investigação sobre os alicerces morais da vida pública moderna. **Revista Sociedade e Estado**, Brasília, v. 33, n. 2, p. 443-464, ago. 2018. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S010269922018000200443&lng=pt&nr=iso>. Acesso em: 11 nov. 2019.

SILVA, M. R.; FUJITA, M. S. L. A prática de indexação: análise da evolução de tendências teóricas e metodológicas. **Transinformação**, Campinas, v. 16, n. 2, p. 133-161, Aug. 2004. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S010337862004000200003&lng=en&nr=iso>. Acesso em: 11 nov. 2019.

STACCIARINI, I. **Polícia entra em confronto com manifestantes contra a Copa do Mundo**. Correio Braziliense. Caderno Cidades. 2014. Disponível em: <https://www.correiobraziliense.com.br/app/noticia/cidades/2014/05/27/interna_cidadesdf,429630/policia-entra-em-confronto-com-manifestantes-contr-a-copa-do-mundo.shtml> Acesso em: 21 out. 2019.

UNISIST. Princípios de indexação. Escola de Biblioteconomia, Universidade Federal de Minas Gerais. Belo Horizonte, v. 10, n.1, p. 83-94, 1981. Disponível em: <<http://www.brapci.inf.br/index.php/article/view/0000002687/81f8286ab6852b2e1872b6c1490ad997>>. Acesso em: 24 set. 2019.

Vocabulário Controlado Básico. Brasília, 2017. **Senado Federal**. Disponível em: <https://www2.senado.leg.br/bdsf/bitstream/handle/id/532112/001106656_VCB_2017_08_20.pdf?sequence=1>. Acesso em: 11 nov. 2019.

WEEDWOOD, B. **História concisa da lingüística**. São Paulo: Parábola, 2002. 165 p. (Na ponta da língua (Parábola) ; 3). ISBN 8588456036.

WITTY, F. J. The Beginnings of Indexing and Abstracting: Some Notes towards a History of Indexing and Abstracting in Antiquity and the Middle Ages. **The indexer**. London, v. 8, n. 4, p. 193-198, 1973. Disponível em: https://www.theindexer.org/files/08-4/08-4_193.pdf. Acesso em: 11 nov. 2019.

ANEXO 1 – Reprodução integral do texto jornalístico analisado

Polícia entra em confronto com manifestantes contra a Copa do Mundo

Isa Stacciarini

postado em 27/05/2014 17:20 / atualizado em 27/05/2014 21:37



(foto: Janine Moraes/CB)

Cerca de 2,5 mil pessoas se reuniram nos arredores do estádio Mané Garrincha em um protesto contra a Copa do Mundo. O grupo saiu da Rodoviária do Plano Piloto por volta das 16h30 desta terça-feira (27/5) e seguiu pela via N1, bloqueando todas as seis faixas, no sentido Palácio do Buriti.

Policiais militares do Batalhão do Choque interditaram o acesso dos manifestantes ao Estádio Nacional, já na altura da Torre de TV. Eles usaram bombas de gás para dispersar os manifestantes, que fazem parte de movimentos como Comitê Popular contra a Copa do Mundo, o Juntos, e do Movimento dos Trabalhadores Sem Teto (MTST). Também estiveram na manifestação os índios que fizeram parte de protestos na Praça dos Três Poderes nesta manhã.



Policial mostra flecha atirada por índio contra tenente da cavalaria, atingido na perna (foto: Kelly Almeida)

Um grupo de manifestantes jogou restos do material da obra das ciclovias contra os militares. Um policial da cavalaria foi atingido por uma flecha na perna, lançada por índios. Segundo a Polícia Civil, duas pessoas foram detidas e encaminhadas à Delegacia de Polícia Especializada (DPE), entre elas, um homem que jogava pedras contra um ônibus.



(foto: Janine Moraes/CB)

Com o acesso ao Eixo Monumental bloqueado pelos manifestantes, alguns motoristas que já estavam na via tiveram que retornar de ré para não entrarem no protesto. Alguns abandonaram os carros e fugiram pelo gramado central. O trânsito no Eixo Monumental provocou reflexos nas pistas do Anexo dos Ministérios. O engarrafamento foi desde o Congresso Nacional até o cruzamento com a W3.

O protesto começou de forma pacífica, com a participação de crianças segurando faixas e cartazes, mas depois o cenário foi mudando e algumas crianças foram vistas correndo, e chorando com seus pais e dentro de ônibus com a mão no rosto, na tentativa de não respirar os gases lançados durante o confronto.

Com informações de Saulo Araújo, Adriana Bernardes, Ariadne Sakkis, Isa Stacciarini, Kelly Almeida, Thais Paranhos e Matheus Teixeira.